IEEE Control Systems
MAGAZINE
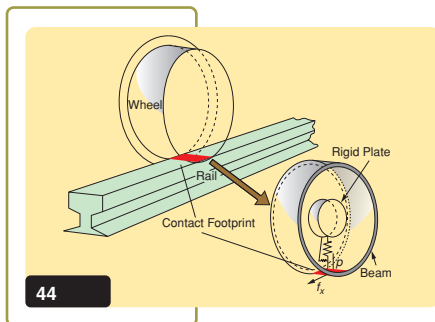
Railway Friction Modeling

IEEE

›› FEATURES

›› DEPARTMENTS

Cover photograph: Courtesy of Jong Kim

21

118

Seeds and Weeds

It's sad to realize that there may very well be a need for a seed vault, a facility that stores thousands of different types of plant seeds so that a species can be reestablished if it becomes extinct in its natural environment. Such a vault now exists. Sitting on a remote Arctic island, the frozen contents of the Svalbard Seed Vault wait silently for the revival that we hope isn't needed.

These seeds are stored to ensure diversity, which is a key element of robustness, the ability to withstand changes in the environment. When all farmers plant the same kind of wheat or the same kind of rice, we know intuitively that we're in trouble. Every system has a weakness, and less diversity means more vulnerability.

Robustness is a central theme in control, but the concept is relevant to virtually every branch of engineering. We know that robustness is attained at the expense of performance. Birds can fly far and fast, but tortoises are better protected. It's tempting to think that shells are more robust than feathers, but these superficial characteristics can be misleading. In fact, any performance/robustness tradeoff depends on how we define each trait—and these attributes can be defined in many different ways.

The human species is extremely robust. We're neither extremely fast nor strong, but we eat a vast range of foods, and we can survive in a wide range of habitats. As a species, we are also diverse, and that accounts for much of our robustness. This robustness makes humans, perhaps, the ultimate "weedy" species—species that manage to invade remote corners of the world and threaten local varieties.

Robustness is strengthened not only by diversity but also by the ability to adapt. In fact, adaptation is possibly the ultimate robustness strategy. While the ability to find a good performance/robustness tradeoff is helpful, a more valuable feature is the ability to develop new robustness and performance traits when the old ones become inadequate.

Since the Svalbard Seed Vault cost a paltry US$9 million to build, I propose a second repository to enhance the chances of the seeds' and thus our own survival. Somewhere in Antarctica might be a good location. After all, traditional robustness wisdom warns against putting all of one's eggs in a single vault.

Dennis S. Bernstein

Dennis visiting NASA Langley in Hampton, Virginia.

MARIO SANTILLO

Dream Topic

Suppose that, at an early age, you were forced to choose a topic in science or engineering that you would spend your entire career on. To be safe, you might search for the "ideal" research topic—a topic that is immediately understandable by everyone, has inestimable value in the commercial sector, is accessible enough to allow modest progress with reasonable effort, involves interesting science and technology ranging from chemistry to mechanics, and is hard enough that it is highly unlikely that someone will

solve the problem and declare the research area dead. This tall order is easily met by the most basic and ubiquitous of physical phenomena: friction.

Friction has two faces. If you wish to avoid slipping and sliding, then increasing friction is the goal. If you wish to increase energy efficiency, then decreasing friction is needed. You can pick either side of the problem, and you'll never run out of challenges. Progress on either side of the problem can have immense commercial value. For control engineers, the problem of controlling motion in the presence of friction is a never ending challenge.

What makes friction so challenging? Like turbulence, friction is a highly nonlinear, emergent phenomenon, that is, a phenomenon with roots in nonlinear dynamics having extremely high dimensionality. Friction models range from atomic-scale physics-based models to macro-scale empirical models. Physics-driven research can use the latest atomic force microscopes to experimentally probe the contributing factors at the smallest scales, while, at larger scales, phenomenological and black box models can be constructed to mimic observed behavior. Small-scale physical insights can be valuable in suggesting

Contributors

Chul-Goo Kang (far right) with his wife Won-Hee, their daughter Da-yeong (a middle school student in Korea), and their son Brian (a junior at Cornell majoring in mechanical engineering).

Brian and Beatrice Armstrong sailing.

Jong Kim visiting the Railway Technical Research Institute in Japan.

Nourdine Aliane on the campus of Universidad Europea de Madrid.

José Luis Guzmán with his wife Aurelia Ramírez in the Odeon of Herodes Atticus at the Acropolis of Athens, Greece.

the structure and parameters of large-scale empirical models, while empirical observations can be used to fine tune models for engineering applications.

Control practitioners must deal with the effects of friction to attain precision motion. Presliding friction—the forces generated before a body moves sizable distances—is also a two-edged sword. Stiction is valuable since it allows a body to remain in place against tiny perturbations, yet it makes motion control challenging since sticking and slipping make it difficult to achieve precise positioning. Consider the volume knob of a radio. A small amount of stiction makes it difficult to make tiny adjustments, yet without any stiction the knob would not remain where you set it. One "solution" to this problem is the use of détentes, as most of us know from a typical car radio. But this quantization solution limits resolution.

In this issue of *IEEE Control Systems Magazine* (*CSM*), we focus on friction

Kris Hollot, his wife Wendy (standing), and daughters Andrea and Erica. Not present for the picture was their son Nicholas.

Jingang Yi and his son Joshua.

modeling and its role in applications. This special section is due to the efforts of Warren Dixon of the University of Florida who serves as guest editor. The first article, by Sung Hwan Park, Jong Shik Kim, Jeong Ju Choi, and Hiro-o Yamazaki, considers the prediction and control of wheel slip in railway vehicles. Unlike rubber automobile tires on asphalt, the contact between metal railway wheels and metal rails entails slipping whenever the brakes are applied. Modeling and controlling this wheel slip is critical to braking performance and longevity.

The second article, by Jingang Yi, provides a detailed analysis of the wafer polishing process, which is critical to semiconductor manufacturing. The importance of this process is clear from the extensive analysis and experimentation that has been devoted to understanding the subtleties of this complex procedure.

The third and fourth articles are more theoretically oriented. The article by Brian Armstrong and Qunyi Chen provides a convenient approach to illustrating the types of physical phenomena captured by various friction models. The last article, by Ashwani Padthe, Bojana Drincic, JinHyoung Oh, Demosthenis Rizos, Spilios Fassois, and myself, has an analogous but narrower goal, where the focus is on the hysteretic properties of various friction models.

For the "Applications of Control" column, Chengliang Liu, Mingjun Wang, and Jun Zhou describe the development of a tractor whose wheels can be independently steered, thus allowing maneuverability that is not achievable by standard vehicles. This department also includes an article by Chul-Goo Kang, who describes the development of a robotic arm that arm wrestles and is kind enough to reward the efforts of its opponents.

Additional highlights of this issue include a response by Kris Hollot to a query about the use of feedback in
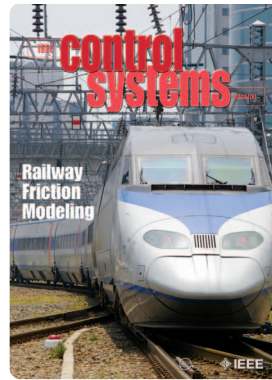
reducing congestion on the Internet, as well as a tutorial article on PID control by J.L. Guzman, K.J. Astrom, S. Dormido, T. Hagglund, M. Berenguel, and Y. Piguet. Along the same lines Nourdine Aliane describes a spreadsheet-based control design and education tool. Education is also highlighted by Malcolm Shuster's article on advice for young researchers, which will surely be of interest to experienced researchers as well.

And that is not all. This issue includes standard columns such as "25 Years Ago," "Technical Committee Activities," "President's Message," "Member Activities," "Bookshelf," and "Conference Reports" as well as an extended "People in Control" column, which includes two interviews along with recognition of 13 2008 IEEE Control Systems Society (CSS) Fellows and six new *CSM* associate and corresponding editors. Congratulations to the Fellows, and welcome to the new board members!

In this issue we also remember Michael Rabins, who left an indelible mark on our field, especially on the enterprise of education.

The issue concludes with a rather unusual application of the CSS logo.

I hope that the excellent contributions to this issue motivate you to contact me about your ideas for future articles for *CSM*. You might have noticed that all *CSM* articles are in full color and, unlike other *CSM* publications, there are no page charges—at any length. This policy recognizes that *CSM* articles provide a valuable service to the CSS community.

Although this issue of *CSM* has no letters, I encourage you to contact me with your comments on any aspect of this magazine, or the control field in general. As CSS President David Castanon likes to say, *CSM* is a conversation about the field. Feel free to speak up!

Dennis S. Bernstein

The Legacy of George S. Axelby

Our Society lost one of its dear friends and pioneers this past June. George S. Axelby died in his home after a prolonged illness. George Axelby was a graduate of the University of Connecticut and Yale and worked for Westinghouse in its Aerospace Division in Baltimore, Maryland. He was the founding editor of *IEEE Transactions on Automatic Control* in 1956 and the founding editor of the IFAC journal *Automatica* in 1968. He was a member of the first group of recipients of the IEEE Control Systems Society Distinguished Member Award in 1983, a recipient of the IFAC Outstanding Service Award, and a Fellow of the IEEE. However, this brief summary barely touches on the impact that George Axelby had on our field.

We have George Axelby's first issue of *IRE Transactions on Automatic Control*, volume 1, issue 1, May 1956, in IEEE *Xplore*. Back then, we were part of the Institute of Radio Engineers (IRE), which merged in 1963 with the smaller American Institute of Electrical Engineers to form the current IEEE. Our Control Systems Society (CSS) was known as the Professional Group on Automatic Control (PGAC), which was formed on October 5, 1954. PGAC was formed as a means to disseminate information about automatic control, its related interests, its problems, and its developments. The charter of PGAC was to organize national group symposia and technical sessions, local professional chapters, and to publish *IRE Transactions on Automatic Control*. This charter has not changed much over the past 54 years.

George Axelby was a member of the administrative committee of the PGAC, corresponding to our current Board of Governors, and he was *Transactions* editor. In his editorial in the first issue, he wrote "Radio and electronic engineers were among those who created new theories and developed more advanced feedback circuits…and a new trend is developing—all of the complex marvels, the radar, the television, the computer, each containing internal feedback control loops, are being combined into over-all automatic control systems." George Axelby foresaw the growth in our field, as he succinctly stated: "In the future automatic control will become more complex; it will encompass broader fields. There will be a greater need for developing and incorporating new theories, techniques, and components into integrated systems."

George Axelby recognized the potential breadth of the field and the role of quality publications in advancing the field. He created standards for paper submission and instituted the foundations of our peer review system, forming an editorial board to evaluate contributions for possible publication. In his words, "the *Transactions* is the medium where this information is distributed; it is the agent where new problems and developments can be proposed." In the early days of the *Transactions*, the edi-

torial board provided the reviews of the submissions. As submissions grew, Axelby organized the editorial board as overseers of our current peer review system, trained the associate editors, and created the information dissemination committee for editorial board meetings where acceptance decisions were made. Axelby retained his position as editor through June of 1968, when he retired to the position of consulting editor.

One of the main reasons for George Axelby's reduced involvement with the *Transactions* was his interest in promoting control in the international control systems community. IFAC was looking to develop an international journal focused on automatic control and was interested in converting the journal *Automatica*, published by Pergamon Press, into an IFAC journal. In 1968, these two organizations approached the one prominent editor in the control field, George Axelby, to become the editor of *Automatica* and convinced him to accept the challenge of developing another quality international publication in the control field. George Axelby soon realized that he had no editorial board or associate editors, so he set out to replicate the editorial structure of the successful *IEEE Transactions on Automatic Control*. He recruited a new international editorial board, one of whom was Huibert Kwakernaak, and set out to find suitable papers. He had the IFAC

symposia papers as potential sources, but they had to be peer reviewed by referees selected by the associate editors. The first issue of the new IFAC journal *Automatica* appeared in January, 1969. To date, *Automatica* remains one of the leading journals in the control field (the latest citation impact factors for controls journals has *Automatica* second after *IEEE Transactions on Automatic Control*.).

George guided *Automatica* through the growth years in our field in the 1970s. By 1979, it was clear that the breadth of systems and control research required an adaptation of the editorial structure. He designed a new editorial structure for *Automatica* to distribute responsibilities by appointing area editors and changing the position of editor to editor-in-chief. A similar structure was eventually adopted by *IEEE Transactions on Automatic Control* nearly a decade later.

George remained as editor-in-chief of *Automatica* until 1993, when interest in *Automatica* increased its publication frequency from six issues per year to monthly. He was recognized by IFAC as editor-in-chief emeritus and was succeeded by Huibert Kwakernaak, who was then deputy editor-in-chief, as the new editor-in-chief. I would be remiss not to mention the significant contributions of Huibert to our Society, as a key member of the editorial staff during George's 25 years and editor-in-chief of *Automatica* for over a decade afterwards. Huibert pioneered the use of informa-

tion technology in support of our publications, as the primary developer of the paper submission and review system Pampus that was used by *Automatica* and as the codeveloper (with Pradeep Misra) of our PaperPlaza conference submission and publication system.

George Axelby's last issue as editor-in-chief of *Automatica* was the 150th issue under his watch. In this issue, he comments on the evolution of *Automatica* from an applications-oriented journal toward a journal of control theory. Of the early issues, he states "there were no theorems, lemmas, or proofs, as most papers do today… there was less concentration on mathematical aspects than there was on creating algorithms for solving practical applications problems." He noted a similar transformation in *IEEE Transactions on Automatic Control*, as he mentions in another of his editorials on the subject of application papers: "it is extremely difficult to define an applications paper particularly in the control field where a mathematical theory is often *applied* to solve a hypothetical control problem." He expected that, as the field matured, more applications papers would be published because control theory would evolve to the point where it should be verified by applications. He foresaw the need for balance in our publications between theory and applications and concluded that "Without eventual application to the physical world, control theory would stagnate and become meaningless."

To George, a good applications paper, "in addition to giving a description of the system and its particular features of interest, should contain the evolution of the control design: its purpose and desired performance, the theory on which it was based, the reasons for choosing the types of mechanization and the problems of realizing it, and finally, some measurements of actual system performance…" These standards are still the guidelines for publications in our *IEEE Transactions on Control Systems Technology* as well as many of our technical articles in *IEEE Control Systems Magazine*.

CSS annually awards the George S. Axelby Outstanding Paper Award to recognize outstanding papers published in *IEEE Transactions on Automatic Control* during the two preceding years. This award was established in 1975 and is one of our most prestigious Society awards. George S. Axelby is also listed in our *IEEE Transactions on Automatic Control* editorial board as the founding editor of the journal.

As our Society nears its 55th year of operation, many of our pioneers are no longer active in our activities or in the control field. Their vision and their efforts led to the creation of our Society and our excellent publications. Our thanks go to George S. Axelby for his crucial and inspirational role in developing the foundations of today's CSS.

David A. Castañón

Worse Is Better

Recall, if you will, the trends of steam engine development as the nineteenth century wore on: increased flyball mass, better-engineered components and consequently reduced friction, faster engines and therefore smaller flywheel moments, striving toward astatic operation and so reduced nonuniformity. Each and all of these trends led away from stable operations. The puzzling increase in hunting behavior was now as clear as day. Simple changes could be made to turn disgruntled and wayward factory engines into well-behaved workers: don't oil the governor, and replace the flyballs with smaller ones. In the long run better engine designs prevailed, and new governors were developed that took account of Vyshnegradskii's words.

—M. Denny, *Ingenium: Five Machines That Changed the World.* Johns Hopkins University Press, Baltimore, MD, 2007, p. 154.
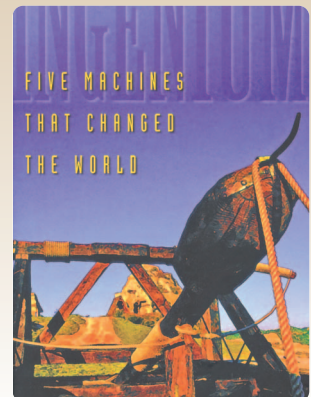
CALL FOR PAPERS
17th MEDITERRANEAN CONFERENCE ON CONTROL AND AUTOMATION

JUNE 24-26, 2009

MAKEDONIA PALACE HOTEL
2, Megalou Aleksandrou Av. – THESSALONIKI, GR-54640
Phone: +30 2310 897197, Fax: +30 2310 897210-11, E. mail: mpalace@grecotel.gr

General Chairs
Vassilios Petridis
Automation & Robotics Lab
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki, Greece
e-mail: petridis@eng.auth.gr

Frank L. Lewis
Automation & Robotics Research Institute
The University of Texas at Arlington
Ft. Worth, Texas, 76118-7115, USA
e-mail: lewis@uta.edu

MCA Founding Chair
Manolis Christodoulou
Dept. of Electronic & Computer Engineering
Technical University of Crete, Greece
e-mail: manolis@ece.tuc.gr

Program Chair
Thomas Parisini
Dept. of Electrical, Electronic, and Computer Eng.
University of Trieste, Italy
e-mail: t.parisini@paperplaza.net

Program Vice Chairs
Zoe Doulgeri
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki, Greece
e-mail: doulgeri@eng.auth.gr

George Rovithakis
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki, Greece
e-mail: robi@eng.auth.gr

Tutorials – Workshops Chair
Thanasis Kehagias
Dept. of Math., Phys., and Comp. Science
Aristotle University of Thessaloniki, Greece
e-mail: kehagiat@auth.gr

Invited Sessions Chair
S. Jagannathan
Dept. of Electrical & Computer Engineering
University of Missouri-Rolla, USA
e-mail: sarangap@umr.edu

Registration & Finance Chair
Loukas Petrou
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki, Greece
e-mail: loukas@eng.auth.gr

Local Arrangements Chair
John Theocharis
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki, Greece
e-mail: theochar@eng.auth.gr

Publication & Publicity Chair
Anastasios Delopoulos
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki, Greece
e-mail: adelo@eng.auth.gr

The 17th Mediterranean Conference on Control and Automation, MED'09, will be held in Thessaloniki, Greece. Thessaloniki, the Cultural Capital of Europe 1997, is a beautiful port city on the Aegean Sea. It is the centre of the Greek North and the country' s second largest city. The conference venue is the seafront Makedonia Palace Hotel. King Cassander of Macedonia founded Thessaloniki in 315 B.C. and named the new city after his wife, Thessaloniki, the sister of Alexander the Great. Throughout its long history, the city has received several designations which are indicative of Thessaloniki's grandeur and glorious past: Capital of Macedonia, Joint-administrative capital of the Byzantine Empire. Thessaloniki has witnessed in the 2300 years of its life the might and glory of the Macedonians, the power of Rome and the grandeur of Byzantium. Its position and the contribution it has made have helped it play a major role in world history. Standing as undeniable proof of that role are its many antiquities, monuments and other impressive archaeological finds, representing every historical period. Within fifty kilometers of Thessaloniki are the beautiful beaches of Halkidiki and several sites of exceptional historical importance. There is Olympus, the mountain of the gods and seat of ancient Greek mythology. Athos, "the Holy Mountain," is a sanctuary for a unique monastic community, safeguarding orthodox Christian tradition and priceless historical treasures. Pella, the ancient capital of the Macedonian Kingdom, features the palace of Alexander the Great, with its exquisite mosaics. Vergina became known worldwide when golden urns containing the remains of Phillip II and his wife - the parents of Alexander the Great - were unearthed during excavation of royal Macedonian tombs. The numerous historical sites, as well as the plethora of natural beauty spots and endless emerald-green shores less than an hour away, establish Thessaloniki as a fascinating journey into art, civilization and relaxation. The conference, through its technical program, will provide a unique opportunity for the academic and industrial community to address new challenges, share solutions and discuss future research directions. A broad range of topics is proposed, following current trends of combining control/systems theory with software/communication technologies. For up-to-date information on MED'09, visit **http://www.med09.org**

TOPICS: Topics of interest include, but are not limited to:

Adaptive control	Industrial automation, manufacturing
Aerospace control	Intelligent control systems
Agents & agent-based systems	Intelligent transportation systems
Approximation-based control	Modeling & simulation
Biologically inspired systems, control	Network control
Computational intelligence	Neural networks
Computing & communications	Non-linear systems
Decentralized control	Optimization
Discrete event systems	Power systems
Distributed systems	Predictive control
Education & training	Process control
Embedded control systems	Real-time control
Fuzzy systems	Robotics
Genetic & evolutionary computation	Robust control
Hybrid systems	Systems biology

IMPORTANT DATES: The Program Chairs are soliciting contributed technical papers for presentation at the Conference and publication in the Conference Proceedings, as well as proposals for invited sessions/papers/talks by topic of interest.

January 15, 2009:	**Contributed, papers, invited session proposals, Workshop / Tutorial proposals, Due**
March 31, 2009:	Notification of Acceptance/Rejection
April 30, 2009:	Final, Camera Ready Papers, Due

PAPER SUBMISSION: Papers must be submitted electronically by January 15, 2009, via the web upload only. The guidelines are given at the MED'09 Web site: **http://www.med09.org** Papers should be submitted in standard IEEE Transactions format. The first page of the paper, centred on the top below the top margin, should include the paper title, the authors' names and their affiliations, an abstract, and keywords. Six pages are allowed for each paper. Up to two additional pages will be permitted for a charge of €125 per additional page. Illustrations and references are included in the page count. Submitted papers will undergo a peer review process, coordinated by the International Program Committee. Authors will be notified of acceptance or rejection by March 31, 2009. Accepted papers in final form must be received no later than April 30, 2009.

Invited Sessions: A summary statement describing the motivation and relevance of the proposed session, invited paper titles and author names must be submitted by the organizer(s) and extended summaries of each prospective invited paper must be submitted by the authors through the conference web site by January 15, 2009.

Workshops-Tutorials: Proposals for workshops - tutorials must be sent by e-mail to the corresponding Chair by **January 15, 2009** and should contain the title of the session, a list of speakers with titles and extended summaries (2000 words) of their presentations.

For more information about the Conference feel free to contact the Program Chairs.

Development and Applications of Multirate Digital Control

DOUGLAS P. GLASSON

This month, "25 Years Ago" revisits the historical background of multirate digital control with Douglas P. Glasson, "Development and Applications of Multirate Digital Control," *IEEE Control Systems Magazine*, vol. 4, no. 4, pp. 2–8, Nov. 1983.

INTRODUCTION

Multirate digital control is a significant area of current research and application that is motivated by practical implementation needs. The motivation for multirate control has traditionally been in aerospace applications where guidance and control laws must be designed to accommodate multiple rates of sensor measurements and finite throughput capabilities of onboard computers. Multirate design techniques should

soon find further utility in control applications for highly distributed systems, such as communication networks, and power-plant/power-distribution networks where the characteristic frequencies and time-constants of a local station's dynamics may differ significantly from those of the network as a whole.

HISTORICAL BACKGROUND

The sheer volume of literature that exists on multirate control techniques underscores the importance of the area and the challenge it presents as a research topic. For example, a literature survey included in a recent paper by Walton [1] notes the contributions of over 50 technical papers related to multirate control techniques. Further, it would be reasonable to suspect that practitioners of digital control have produced an immense body of unpublished "methods that work" in

developing and implementing practical systems. In this paper, the origins and evolution of multirate control techniques are summarized briefly, noting those contributions that have provided the basis for the more popular techniques now in use. A historical overview of digital control development is presented in Fig. 1 (Figure 1 was adapted from a similar chart presented in a thesis by Amit [2]). The field of digital control, or more precisely, sampled data control, originated in radar applications during World War II. Because the rotating antenna of a radar system illuminates a target only intermittently, early radar-aided tracking and fire-control systems had to be designed to utilize data in sampled form. Methods for effective design of control systems using sampled data were under initial development during the later 1940's, and multirate

systems theory followed these efforts in the early 1950's.

Initially, researchers developed multirate techniques as a method of evaluating more conventional types of controllers such as continuous systems and single-rate sampled data systems. For example, one could study the inter-sample behavior of a signal or output of a single rate control system by introducing a "phantom sampler" (i.e., a fictitious sampler that operates at a rate some integer ratio higher than that of the controller). A significant early contribution to this general method of analysis, known as frequency decomposition, was made by Sklansky and Ragazzini [3] who described the use of this technique in error-sampled control system development. In the late 1950's, Ragazzini and Franklin [4] published a textbook that described both this technique and the closely related switch decomposition technique. Friedland [5] later related the frequency decomposition technique to periodically varying control structures, followed by contributions of Coffey and Williams and Boykin and Frazier which dealt with the analysis

of multiloop, multirate control structures (multiloop referring to a feedback control structure having nested single-input/single-output compensating elements).

Shortly following the origin of the frequency decomposition technique, a similar frequency domain technique known as switch decomposition was developed. Researchers had begun to see the potential value of multirate systems beyond being a technique for analyzing single-rate systems; switch decomposition seemed a "natural" approach to developing such systems. The switch-decomposition technique attributed to Kranc [6], provided a means of representing a multirate control structure as an equivalent single-rate controller; this representation accomplished, the controller could be designed and analyzed using existing singlerate techniques. In the late 1960's, Jury [7] showed an equivalence of the switch decomposition technique and the frequency decomposition technique. Recently, Whitbeck has developed a vector form of the switch decomposition technique and applied it to various problems in flight control [8–10].

Time-domain methods of multirate stability analysis and design were initiated by Kalman and Bertram [11] with the publication of their state space stability analysis technique in 1959. This paper made a major contribution in showing the power and flexibility of state space techniques in characterizing many types of sampled data control systems, including time-varying systems. Apparently, little significant work was initiated to build on this work for nearly fifteen years. Barry [12] published a paper in 1975 in which he described the design of a multirate regulator and showed that its performance was superior to a single-rate regulator having the same base (slow) sample rate. During 1979–81, researchers at The Analytic Sciences Corporation (TASC) developed a new multirate control design technique based on an optimal estimation and control formulation. This research [13–16] included mathematical formulation of the design problem, development of computational design techniques, and applications of these techniques to flight control examples. Essentially in parallel with the work at TASC, Amit and Powell [2] independently investigated a similar
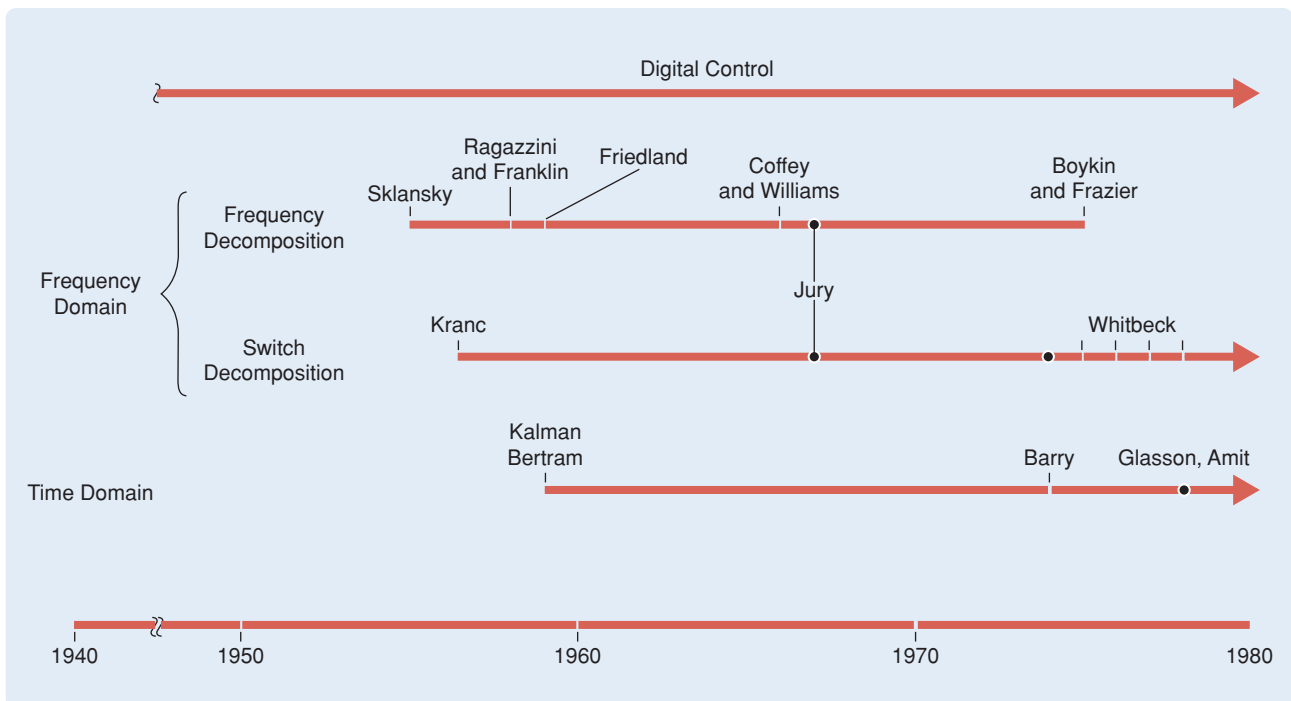
FIGURE 1 Development of multirate digital control.

optimal control formulation; their work resulted in, among other things, some practical considerations for implementing multirate control laws and a highly efficient method for solving periodic Riccati equations related to the design techniques.

REFERENCES

[1] V.M. Walton, "State–space stability analysis of multirate–multiloop sampled data systems," *Proc. of the AAS/AIAA Astrodynamics Specialist Conference*, Lake Tahoe, NV, August 1981.

[2] N. Amit, "Optimal Control of Multirate Digital Control Systems," *Report No. SUDAAR #523, Stanford University*, Stanford, CA, July 1980.

[3] J. Sklansky and J. R. Ragazzini, "Analysis of errors in sampled-data feedback systems," *AIEE Trans.*, vol. 74, part II, pp. 65–71, May 1955.

[4] J.R. Ragazzini and G. F. Franklin, *Sampled Data Cortrol Systems, John Wiley and Sons, Inc.*, New York, 1958.

[5] B. Freidland, "Sampled-data control systems containing periodically varying members," *Proc. of the 1st International Federation of Automatic Control*, Moscow, May 1959.

[6] G.M. Kranc, "The Analysis of Multiple-Rate Sampled Systems," *Report No. T-11/B, Electronics Research Laboratory, Dept. of Electrical Engineering*, Columbia University, New York, NY, May 1956.

[7] E.I. Jury, *Sampled Data Control Systems, John Wiley and Sons, Inc.*, New York, 1958.

[8] R. F. Whitbeck and L. G. Hofmann. "Analysis of Digital Flight Control Systems with Flying Qualities Applications," *Report No. AFFDL-TR-78-115*, vol. II, September 1978.

[9] R.F. Whitbeck and L.G. Hofmann, "Digital control law synthesis in the w' domain," *AIAA J. Guidance and Control*, vol. 1, no. 5, pp. 319–326, September–October 1978.

[10] R.F. Whitbeck and D.G.J. Didaleusky, "Multirate Digital Control Systems with Simulation Applications," *Report No. AFWAL-TR-80-3101*, vol. I, II, III. September 1980.

[11] R.E. Kalman and J. Bertram, "A unified approach to the theory of sampling systems," *J. Franklin Institute*, vol. 267, pp. 405–436, May 1959.

[12] P.E. Barry, "Optimal Control of Multirate Systems," *Report No. RN-361, Research Department*, Grumman Aerospace Corporation, July 1975.

[13] D.P. Glasson, "Research in Multirate Estimation and Control," *Report No. TR-1356-1, The Analytic Sciences Corporation*, Reading, MA, December 1980.

[14] D.P. Glasson and J. Dowd, "Research in Multirate Estimation and Control-Optimal Sample Rate Selection," *Report No. TR-1356-2, The Analytic Sciences Corporation,* Reading, MA, October 1981.

[15] J.R. Broussard and D.P. Glasson, "Optimal multirate flight control design," *Proc. of the 1980 Joint Automatic Control Conference*, San Francisco, CA, August 1980.

[16] D.P. Glasson, "Robustness Properties of a New Multirate Digital Control System," *Presented at AIAA Aerospace Sciences Meeting*, St. Louis, MO, January 1981.

Contributed by Kent Lundberg
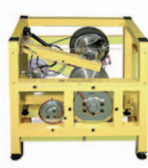
In this issue of *IEEE Control Systems Magazine*, we invite Kris Hollot of the University of Massachusetts to respond to a query on the use of feedback control to keep the Internet running smoothly. Readers are invited to submit technical questions, which will be directed to experts in the field. Please write to us about any topic, problem, or question relating to control-system technology.

WORLDWIDE FEEDBACK

Q. **In my computer science course, the professor mentioned that control is used to help make the Internet run smoothly. Can you please explain to me how that is possible?**

Kris: This is a great question! I'm delighted to have the chance to answer it.

The Internet is a digital information highway connecting individual computers with computer servers acting as data receivers and data senders. When the Internet is running smoothly, data encounter little congestion and, upon arrival at the intended receiver, trigger an acknowledgment back to the data-sender signaling successful transmission. Feedback control mechanisms play important roles in keeping this highway running smoothly.

The Internet, and the data it handles, is not too different from the interstate highway system and the car and truck traffic it handles. Digital data pipes play the role of highways, data represent automobiles, and Internet routers, which direct data from one data pipe to another, represent interstate highway entrance and exit ramps. One difference between the two is that the Internet actively manages the rates at which data are injected into the network and actively works to facilitate reliable data file

transfer. The Internet does this through a set of rules referred to as the Transmission Control Protocol (TCP), which relies on feedback signals generated within the Internet.

To describe these feedback signals, let's consider sending data from a computer in Boston to another computer, say, in Los Angeles (LA). By way of the automobile-traffic analogy, consider driving a car from Boston to LA. The automobile encounters other traffic as it traverses the interstate highways; it may encounter some congestion on entrance and exit ramps, and, if all goes well, it will eventually reach LA, at which time a call-back to Boston acknowledges that the trip was successfully concluded.

The situation with the Internet is similar. If a piece of data successfully reaches the LA computer (having traversed several transcontinental data pipes), another piece of data, called an acknowledgment (*ack*), is sent back to the Boston computer, signaling successful data transmission. This ack makes its way back to the Boston computer using the reverse Internet path.

The ack I've just described is more accurately referred to as a *positive* ack, to distinguish it from a *negative* ack, which is a different feedback signal indicating that the piece of data has not reached its intended destination. In both cases, acks are Internet feedback signals that provide real-time

status of the Internet's ability to deliver data.

The fact that a piece of data might not reach its Internet destination may be somewhat surprising. But data can be "dropped" in the Internet when data rates persistently exceed the data-carrying capacity of data pipes. When this happens, the Internet's entrance and exit ramps become jammed up, similar to congested highways. In the Internet, these entrance and exit ramps are data buffers (memory), and, when they are completely filled, the next entering data piece is literally dropped and thus destroyed—kaput! Although the dropped data can't make it to its destination, the Internet must ascertain that the dropped data did not reach their receiver. Having done so, the Internet requests that the data sender retransmit the dropped piece of data.

Now let's see how the Internet uses these feedback signals to take actions necessary to keep the system running smoothly. In control engineering language—let's see how we can use this sensor output to actuate the system to meet some control objective.

Rather than sending just a single piece of data, let's now consider transmission of a large data set across the Internet (think large image files). Now TCP comes into play. Roughly speaking, one function of TCP, specifically, its sending-rate protocol, determines the rate at which data senders inject data into the Internet. Moreover, this protocol reacts to positive and negative acks and in this sense is a feedback controller. In fact, every computer has such a feedback controller to modulate the rate at which it injects data into the Internet. A

feedback control system description of the Internet now begins to emerge. The sending-rate protocol, in reaction to feedback signals, dictates rates at which data are injected into the Internet. In turn, these data fill the Internet data pipes and possibly contribute to congestion. Finally, as data are received or dropped, positive and negative acks are generated, and a new round of feedback signals is delivered to the sending-rate protocol; see Figure 1.

Now, here's the interesting question for control folks: just how are these acks used to modulate data-sending rates?

Similar to many feedback controllers, the actions taken by the sending-rate protocols are intuitive: if a sender's data are deemed to congest the Internet, then its data rate is decreased. Conversely, if a sender's data are not a congesting factor, then the sending-rate increases.

The Internet decides whether a particular sender's data are congesting by using the acks as a congestion signal. Dropped data indicates that a source contributes to overall Internet congestion. Consequently, if a data sender receives a negative ack, the protocol decreases the data rate, whereas, if a positive ack is received, data rates increase.

The amount and quality of rate increase or decrease is important and contributes to transient behavior and the value of steady-state sending
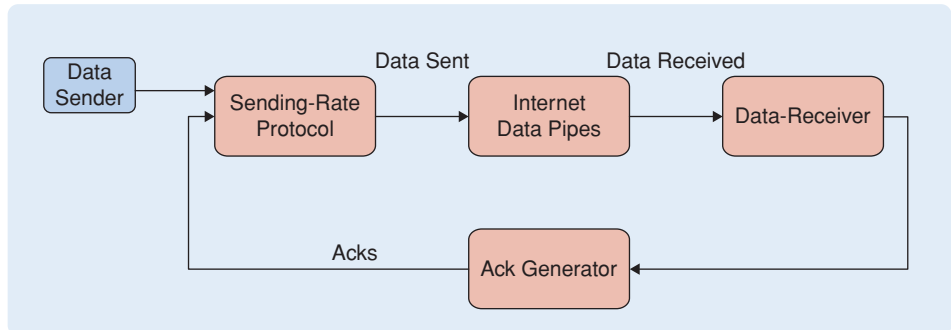
FIGURE 1 This block diagram depicts the Internet's sending-rate protocol as a feedback controller. The sending-rate protocol injects data into the Internet's data pipes. Congestion may occur in these pipes and cause data to be dropped. As data is received (or dropped), positive and negative acks are generated, and a new round of feedback signals are delivered to the sending-rate protocol. If the data sender receives a negative ack, the protocol decreases the data rate, whereas, if a positive ack is received, data rates increase. Thus, the Internet reacts, in negative feedback fashion, to congestion.
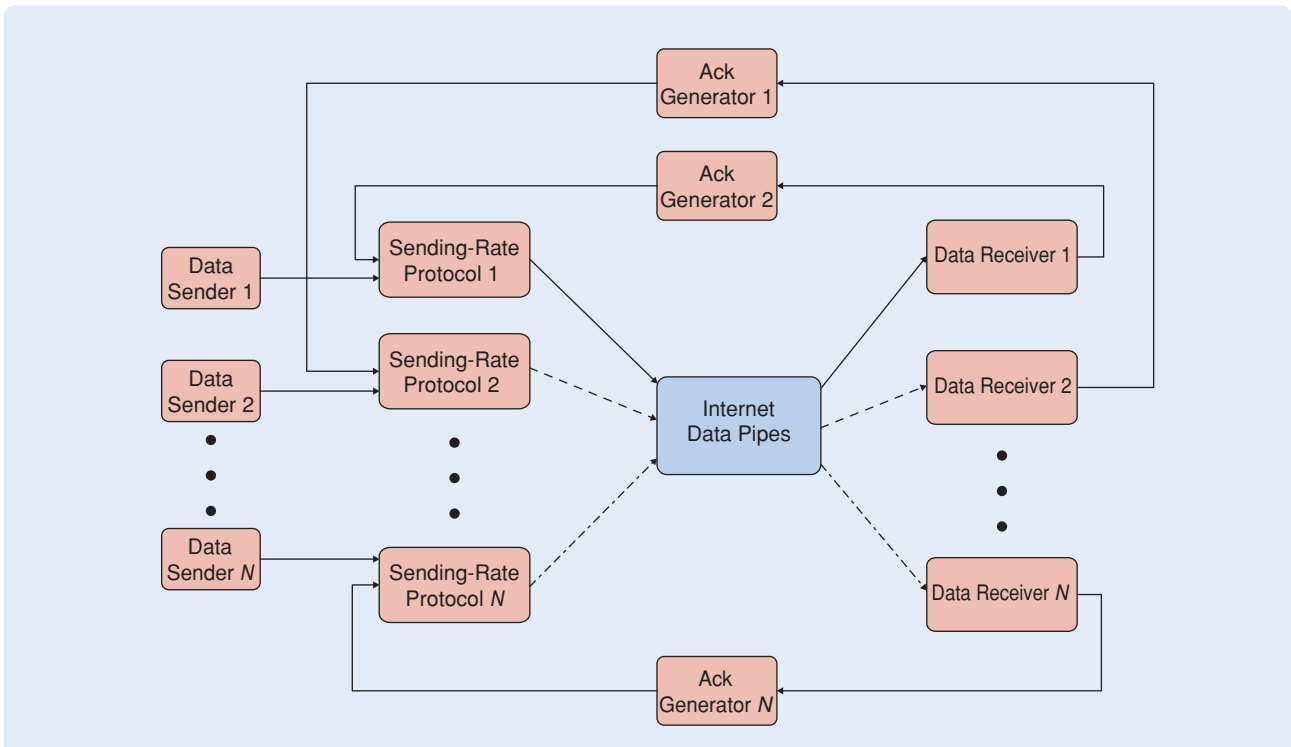
FIGURE 2 This block diagram generalizes Figure 1 by allowing for the realistic case of many data senders and data receivers. The Internet data pipes are a multi-input, multi-output set of fully interconnected pathways. The feedback control structure is decentralized wherein the sending-rate protocol for a specific sender depends only on the acks generated by its data. Thus, a data sender reacts to congestion experienced by its own data but does not react to overall Internet congestion. For simplicity, we have ignored the fact that the acks make their way ack to the data senders by means of the Internet data pipes.

the M.S. and Ph.D. in electrical engineering from the Ohio State University in 1991 and 1995, respectively. From 1995 to 2004 he was a faculty member at the U.S. Naval Academy. Since 2004 he has been an associate professor in the Department of Electrical and Computer Engineering at Wichita State University. He served as vice-chair for interactive papers of the 2007 CDC, registration chair of the 2006 ACC, and finance chair of the 2002 ACC. Presently, he is Student Affairs chair for the 2008 IEEE Multiconference on Systems and Control. He has been an associate editor of the Conference Editorial Board of the IEEE Control Systems Society since 1999. He was a NASA-ASEE Summer Faculty Fellow in 1999 and 2000. In 2007 he received the Wallace Excellence in Teaching Award. His research interests include time-delay systems, fault detection and isolation, active magnetic bearings, active noise control, mobile sensor networks, and spacecraft dynamics and control. (CSS)

---

rates. One TCP variant, TCP-Reno, increments sending rates by a fixed amount for each positive ack, and approximately halves the rate on receipt of a negative ack. Said another way, we have the control law

$$\frac{d}{dt} \text{ sending rate}$$
$$= \begin{cases} \text{constant}, & \text{if positive ack} \\ -\frac{1}{2} \text{ sending rate}, & \text{if negative ack.} \end{cases}$$

Consequently, the amount of data injected into the Internet increases linearly until congestion is sensed (by means of a negative ack) and then backs off exponentially.

Up to this point, we have presented a simplistic view of feedback control mechanisms in the Internet. In reality, the situation is much more complex, providing challenges to systems modeling, dynamical analysis, and control design; see Figure 2.

As a dynamic system, the Internet is an example of a hybrid system wherein actions are triggered by events; for example, the decrease in sending rates is triggered by congestion events. The Internet is also an example of a large-scale decentralized control system, where many data senders inject data into Internet datapipes, each bound for a specific datareceiver; where each sending-rate protocol reacts only to congestion signals generated by its data.

The existence of time delays in the Internet also plays a crucial role in feedback control, a situation that every student of feedback control can appreciate. These time delays are comprised of the propagation delays associated with data traveling through data pipes at the finite speed of light as well as by queuing delays encountered by data moving through the Internet's congested entrance and exit ramps. As in any feedback control system, these delays restrict the gain of a feedback controller or, in the case of the Internet, restrict the aggressiveness with which the sending-rate protocols react to congestion signals. Finally, the whole notion of "controller implementation" or "sensor implementation" in the Internet is challenging, since it must be accomplished using the data itself and existing protocol structures.

In this discussion, I've given a high-level view of the role that feedback plays in keeping the Internet running smoothly. Many important details have been abstracted out. To probe further, we recommend the three sources below including the references contained therein.

REFERENCES

[1] J.F. Kurose and K.W. Ross, *Computer Networking: A Top-Down Approach*, 4th ed. Reading, MA: Addison-Wesley, 2008.
[2] R. Srikant, *The Mathematics of Internet Congestion Control*. New York: Birkhäuser, 2003.
[3] S.H. Low, F. Paganini, and J.C. Doyle, "Internet congestion control," *IEEE Contr. Syst. Mag.*, vol. 22, no. 1, pp. 28–43, Feb. 2002.

AUTHOR INFORMATION

Kris Hollot received the Ph.D. in electrical engineering from the University of Rochester, New York, in 1984. Since 1984, he has been with the Department of Electrical and Computer Engineering at the University of Massachusetts, Amherst. He has served as associate editor for several control journals, and his research interests are in the theory and application of feedback control. He received the National Science Foundation Presidential Young Investigator Award in 1988. (CSS)

Developing Countries Conference Support Program

The new IEEE Control Systems Society (CSS) Developing Countries Conference Support (DCCS) Program will run as a pilot program during 2008–2010. The purpose of DCCS is to promote research in control systems engineering and theory in developing countries by supporting, intellectually and financially, conferences that take place in these countries. To be eligible, a conference for which the support is requested must be held in a country that belongs to the list of developing countries (at the time the request is made) and must receive CSS sponsorship or technical cosponsorship. A list of developing countries is provided by the World Bank, at http://web.worldbank.org. Evaluation is based on quality and on the conference organizers' plans for outreach and impact in developing nations. Support provided may not exceed US\$20,000 for each conference and may not exceed 50% of the budgeted conference expenditure.

Supported conferences must acknowledge financial support of the CSS. The DCCS budget is US\$40,000 per year.

Proposals are considered twice each year. The deadlines for submissions are May 1 and November 1. Notifications of awards are made by June 30 and December 31 each year.

Support can be sought for the following:

» Travel and lodging expenses for plenary speakers. Extended biographies must be provided for the proposed speakers.
» Student/newcomers reception, where CSS membership forms will be made available to all attendees.
» Funds to defray costs of registration fees for students or for CSS members.
» Funds to defray costs of registration fees for people from developing countries.
» Expenses for technical workshops associated with the conference.

Proposals may also seek support for other conference activities not mentioned above; the applicants should describe in detail the scientific merits and benefits to conference participants and to the field. Proposals must include a detailed budget for the conference as well as a detailed justification for the requested funding. Within six months of the conclusion of the conference, a report on the use of the funding must be submitted to the CSS vice president for member activities.

The evaluation committee for this year consists of Elena Valcher, Rick Middleton, Frank Allgöwer, and myself. Please contact me at tomlin@eecs.berkeley.edu for more information, or visit our Web page, linked to www.ieeecss.org, for an online application.

Special thanks go to Elena Valcher and Rick Middleton for proposing and designing this program and for taking the steps to have it approved. I look forward to receiving your applications.

Claire Tomlin

Hired Gun

[M]axwell] was the first person to understand mathematically how Watt governors work. He was asked to look into the problem because of its pressing nature, and because of his known mathematical abilities. It may seem like something of a comedown for the man who analyzed Saturn's rings, and showed why they are stable, to turn his gaze from the heavens to an oily mechanical device that was troubling factory workers. In fact, however, Maxwell was able to use the same mathematical technique in both cases…. He reduced complicated equations that were impossible to solve exactly to simple equations that were easily solved and that yielded approximate solutions to the original problem. In the case of the Watt governor, these simplified equations told him the answer to the hunting problem.

—M. Denny, *Ingenium: Five Machines That Changed the World*, Johns Hopkins University Press, Baltimore, MD, 2007, p. 149.
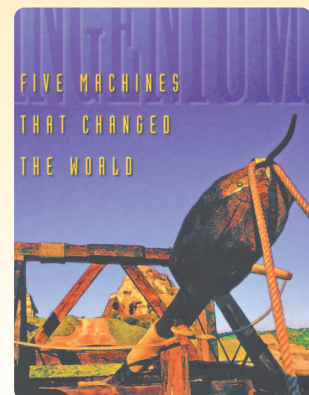
Technical Committee on Intelligent Control

This column focuses on the activities of selected technical committees (TCs). In addition to providing information about TC activities, this column is intended to serve as a call for participation by readers with related interests. Please contact Jay Farrell at farrell@ee.ucr.edu for information about joining an IEEE Control Systems Society TC.

The IEEE Control Systems Society (CSS) Technical Activities Board met at the 2008 American Control Conference in Seattle on Wednesday June 11, 2008. In attendance were Richard Braatz, Michael Demetriou, Jay Farrell, Shuzhi Sam Ge, Mustafa Khammash, Daniel Rivera, and Andy Teel. One of the main topics of discussion was the role of the technical committees (TCs) in organizing the technical program of the various CSS conferences. In particular, the TCs have played a significant role in motivating, creating, and organizing the Multiconference on Systems and Control (MSC).

The IEEE MSC provides an organizational structure for the MSC hotel, registration, and financial aspects allowing each individual conference program committee to concentrate on technical excellence in the papers and workshops. The MSC offers TCs the infrastructure within which to rapidly organize meetings, workshops, symposia, and conferences on emerging and existing topics. Registrants to the IEEE MSC can attend and participate in activities within each subconference, paving the way for greater scientific exchange, exposure to new disciplines, and increased interaction among participants in one centralized location. Under the overarching umbrella of the multiconference, attendees are provided with a consolidated proceedings volume.

The IEEE Conference on Decision and Control (CDC) and the IEEE MSC are the two major conferences organized, sponsored, and governed by the CSS. The first MSC, which was held in 2007, combined the IEEE Conference on Control Applications (CCA) and the IEEE International Symposium on Intelligent Control (ISIC). In 2008, MSC also included the IEEE International Symposium on Computer-Aided Control System Design (CACSD). This article focuses on the role of the Technical Committee on Intelligent Control (TCIC) in organizing these conferences.

INTELLIGENT CONTROL FOCUS

The objectives of TCIC are to make fundamental research contributions in intelligent control, to promote intelligent system development in providing solutions for increasingly complex industrial systems, and to facilitate scientific research and exchange in a vibrant and collegial environment. To achieve these objectives, TCIC is dedicated to addressing the fundamental issues of the present and the future in intelligent control and is organized into ten working groups: 1) adaptive control, 2) distributed control systems, 3) fuzzy systems, 4) neural networks, 5) learning techniques, 6) knowledge-based systems, 7) hybrid intelligent systems, 8) soft computing, 9) architectures of intelligent control systems, and 10) applications of intelligent control systems.

The increasing complexity and sophistication of modern systems calls for intelligent control systems to ensure higher performance in diverse operating conditions, some of which may fall outside of the scope of conventional control. Intelligent control involves the seamless fusion of systems and control, computer science, and operational research, among others, coming together, merging, and expanding in new directions and opening new horizons to address the problems of this challenging and promising area. Intelligent control systems are typically able to perform one or more of the following functions: planning actions at multiple levels of detail, emulation of human expert behavior, learning from past experiences, integrating sensor information, identifying changes, and reacting appropriately.

TCIC members are active in the control research community as indicated by the monographs, journal, and conference publications listed on the TC Web site. In addition to publications, TCIC members also contribute to the control community by serving on control conference committees, serving as editors and associate editors of control and computational intelligence related journals, and leadership of the CSS.

TCIC continues to play an active role in organizing workshops, symposia with international academic organizations, participating in IEEE conferences in areas pertaining to intelligent control system research and design, coordinating academic and

TABLE 1 Summary of ISIC organizational data.

General Chair	Program Chair	Location	Dates
G.N. Saridis	A. Meystel	Troy, New York	August 26, 1985
A. Meystel	J.Y.S. Luh	Philadelphia, Pennsylvania	January 19–20, 1987
H.E. Stephanou	A. Meystel, J.Y.S. Luh	Arlington, VA	August 24–26, 1988
A.C. Sanderson	A.A. Desrochers, K. Valavanis	Albany, New York	September 25–26, 1989
A. Meystel	H. Kwatny, S. Navathe, H. Wechsler	Philadelphia, Pennsylvania	September 5–7, 1990
H.E. Stephanou	A.H. Levis	Arlington, Virginia	August 13–15, 1991
E. Grant	T.C. Henderson	Glasgow, Scotland	August 11–13, 1992
P. Antsaklis	K.M. Passino, U. Ozguner	Chicago, Illinois	August 25–27, 1993
U. Ozguner	L. Acar, M.B. Leahy, Jr.	Columbus, Ohio	August 16–18, 1994
K. Valavanis	L. Acar, M.B. Leahy, Jr.	Monterey, California	August 27–29, 1995
K.M. Passino	J.A. Farrell	Dearborn, Michigan	September 15–18, 1996
Y.I. Stefanopoulos/	K. Ciliz/E. Barbier, U. Ozguner	Istanbul, Turkey	July 16–18, 1997
J. Albus	S. Lee, A. Koivo, A. Meystel, F.-Y. Wang	Gaithersburg, Maryland	September 14–17, 1998
M. Kokar	M. Lemmon	Cambridge, Massachusetts	September 15–17, 1999
P. Groumpos	M. Polycarpou	Patras, Greece	July 17–19, 2000
Daniel Repperger	Thomas Parisini	Mexico City, Mexico	September 5–7, 2001
C. W. De Silva	Fakhri Karray	Vancouver, British Columbia, Canada	October 27–30, 2002
G.G. Yen	S. S. Ge	Houston, Texas	October 5–8, 2003
S.S. Ge	T. Samad	Taipei, Taiwan	September 2–4, 2004
M.M. Polycarpou	T. Parisini	Limassol, Cyprus	June 27–29, 2005
Andras Varga	Derong Liu	Munich, Germany	October 4–6, 2006
Shuzhi Sam Ge (MSC)	Sarangapani Jagannathan	Singapore	October 1–3, 2007
Oscar Gonzalez (MSC)	Kevin Moore	San Antonio, Texas	September 3–5, 2008

research activities, and providing relevant information to interested researchers. Members of TCIC first organized the IEEE International Symposium on Intelligent Control (ISIC) in 1985. The committee now has a 22-year history of successful operations as summarized in Table 1. This symposium has been held jointly with the CCA or CACSD in 1996, 2001, 2004, and 2006. In 2000 and 2005, ISIC was held jointly with the Mediterranean Conference on Control and Automation. Since 2007, it has been organized under the MSC. At various times, the ISIC has been jointly organized with IFAC and one or more of the IEEE Computer, Industrial Electronics, and Robotics and Automation Societies.

TCIC is dedicated to providing informational forums, conferences for technical discussion, and information over the Web to researchers in the CSS who are interested in the field of intelligent control and its applications. The TCIC newsletter was set up in 2007 to serve this purpose. CSS members who are interested in participating in TCIC are welcome to contact the leader of the working group nearest to their interests. TCIC is currently chaired by Shuzhi Sam Ge. The TC Web site can be found through the CSS Web site or at http://robotics.nus.edu.sg/tcic/.

Jay Farrell

Better Is Worse

The engineers of 1868 were puzzled and worried that the unruliness of their steam engines seemed to be getting worse. Looking back to Watt's time and the intervening decades, they saw that the problems were increasing as engines became better engineered, which seemed counterintuitive to them, and seems counterintuitive to those of us today who are not control engineers. The search for astatic performance failed. Truly astatic engines could not be built, and engines that approximated this state—those with reduced dependence of engine speed on load—behaved worse than those with a marked dependence, despite the known offset problem. How could the small, crude engines of the earlier days, with a simple governor borrowed from windmills, behave better than the sleek, carefully manufactured behemoths of 1868?

—M. Denny, *Ingenium: Five Machines That Changed the World*, Johns Hopkins University Press, Baltimore, MD, 2007, pp. 148–149.
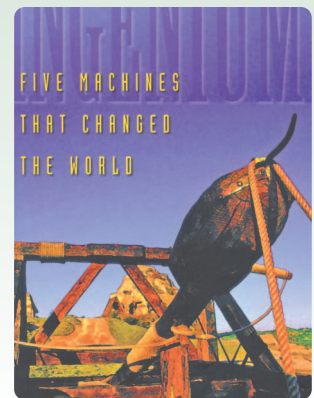
Coordinating Control for an Agricultural Vehicle with Individual Wheel Speeds and Steering Angles

CHENGLIANG LIU, MINGJUN WANG, and JUN ZHOU

Automatic navigation of agricultural vehicles, such as tractors and harvesters, can increase the level of automation in the agricultural process, thereby lessening the human workload [1], [2]. In the agricultural environment, where vehicles move on uneven terrain, achieving maneuverability and mobility can be difficult, particularly when the path curvature is small or vehicle orientation and path tracking must be independently controlled.

To obtain the desired maneuverability and mobility, one approach is to individually actuate the drive and steering functions of each wheel on a four-wheeled vehicle. Figure 1 shows a prototype system for agricultural purposes. Figure 2 shows a vehicle schematic and block diagram model for closed-loop control of the steering and driving motors. For sensing and navigation, this prototype vehicle uses a combination of machine vision, a differential global positioning system (DGPS), and an inertial measurement unit (IMU). The focus of this work, however, is on path-following control. Although the vehicle is highly maneuverable due to its eight independent actuators, precise control is difficult due to the overconstrained nature of the actuation. It is challenging to coordinate all of the motors to avoid wheel slippage [3] as well as to precisely control the vehicle's motion in both translation and orientation.

DETERMINATION OF CONTROL INPUTS

Analysis of the rigid-body velocity is used to determine the instantaneous control values for all of the motors. Assuming that the vehicle is a rigid body, the absolute velocity vector for each wheel is the sum of the translational and rotational velocity vectors with respect to the geometric center. The vehicle has

FIGURE 1 An eight-degree-of-freedom autonomous agricultural navigation platform. Each wheel is equipped with a driving motor and a steering motor. Thus, each wheel can be individually steered and driven.
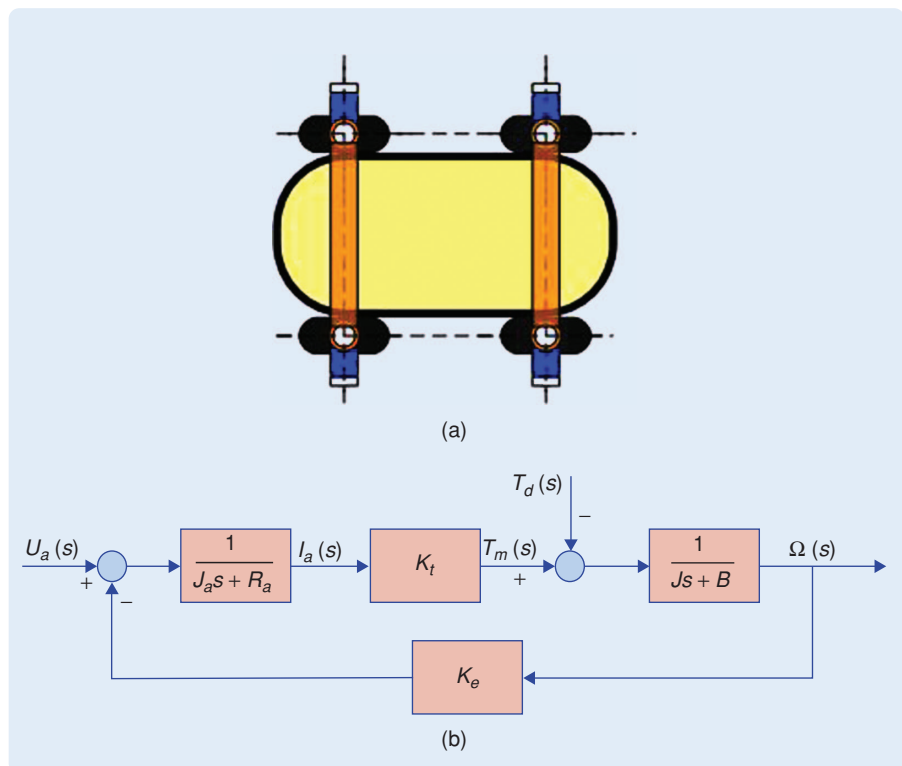
FIGURE 2 Vehicle schematic. (a) The vehicle has four wheels, which are individually steered and driven. (b) A model of the driving and steering motors. The model can be reduced to an inertia link in cases of small armature inductance.
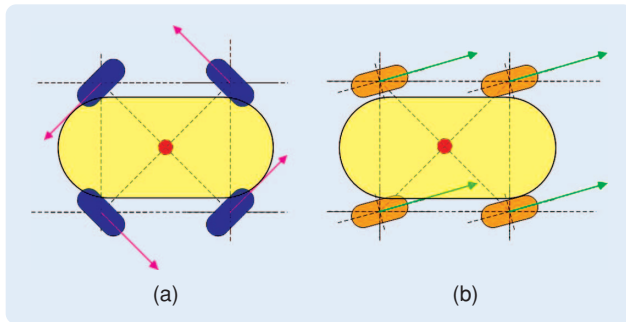
FIGURE 3 Rotational and translational motion. (a) Rotational motion. Each wheel is oriented 45° or −45° with respect to the vehicle body, and all four wheels spin at the same rate. The vehicle body thus rotates around its geometric center at a fixed rate with zero turning radius. (b) Translational motion. Each wheel is oriented at the same angle with respect to vehicle body and the global reference frame, and all four wheels spin at the same rate. Thus, the vehicle body translates in the global reference frame without a change of attitude.
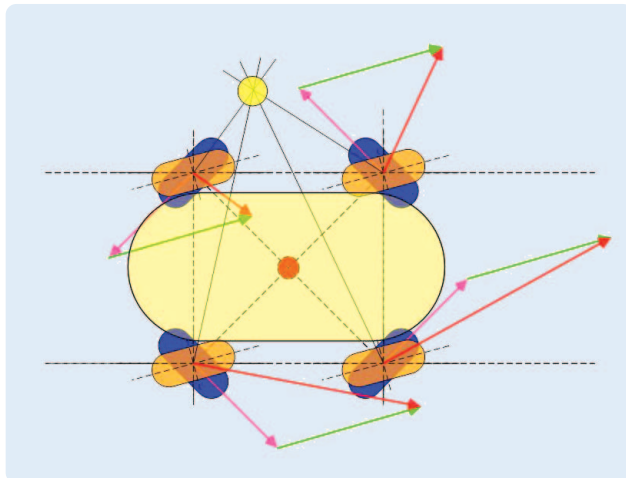
FIGURE 4 Velocity decomposition for all wheels. For each wheel, the pink vector represents the rotational component, the green vector describes the translational component, and the red vector is the composition of these velocities as a result of the parallelogram rule.
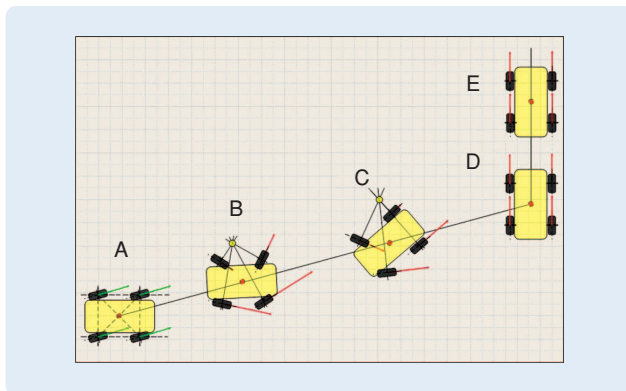
FIGURE 5 Vehicle motion along the reference path (ABCDE). The geometric center, represented by the red dot, tracks the reference path exactly. The attitude of the vehicle changes gradually throughout the path-following task.

four wheels that are individually driven and steered, thereby making it omnidirectional. Unlike other agricultural navigation platforms, the vehicle can perform motions that are either purely translational or purely rotational. We first analyze the translational velocity and rotational velocity vector for each wheel as shown in Figure 3 and then study the composition of these velocity vectors as shown in Figure 4.

During purely translational motion, the longitudinal directions of all four wheels are oriented identically with respect to the vehicle body, and all four wheels spin at the same rate around their drive axes. Therefore, as shown in Figure 3(b), all of the wheel centers, and thus the geometric center of the vehicle body, move at the same velocity.

For purely rotational movement, the longitudinal axis of each wheel is oriented at either 45° or −45°with respect to the orientation of the vehicle body, as shown in Figure 3(a). Due to the geometry of the vehicle, the velocity orientation of each wheel center is perpendicular to the line between the vehicle geometric center and the respective wheel center. Since all of the wheels are commanded to rotate at the same angular velocity, the vehicle rotates as a rigid body around its geometric center. Consequently, the vehicle's turning radius is zero, giving the vehicle maneuverability that is superior to most two-wheel-steered agricultural vehicles.

ANALYSIS

We now combine the individual wheel translational and rotational motions to obtain the orientation and velocity of the vehicle as a whole. Figure 4 shows the translation (green) and rotational velocity vector (red) of each wheel center. Additionally, illustrated by the red vectors is the sum of the translational and rotational velocity vectors for each wheel. For each wheel, it follows that

$$\mathbf{u}_c = \mathbf{u}_t + \mathbf{u}_r, \qquad (1)$$

where $\mathbf{u}_c, \mathbf{u}_t$, and $\mathbf{u}_r$ denote the total velocity vector, translational velocity vector, and rotational velocity vector of the wheel, respectively. Assuming no slip, the translational velocity of each wheel equals that of the vehicle, both of which are denoted by $\mathbf{u}_t$. In addition, $\mathbf{u}_r$ is given by

$$\mathbf{u}_r = \mathbf{r} \times \boldsymbol{\omega}, \qquad (2)$$

where $\mathbf{r}$ is the vector from the instantaneous center of rotation (ICR), illustrated by the yellow circle of the vehicle in Figure 4, to the wheel itself, and $\boldsymbol{\omega}$ is the angular velocity of rotation for the vehicle. Here, $\beta_r(t)$ represents the angle of the vector $\mathbf{u}_c$ with respect to the x-axis of a global reference frame, and $\omega_r(t)$ denotes the length of the vector $\mathbf{u}_c$.

The control inputs for the steering and driving motors $U_{as}(t)$ and $U_{ad}(t)$ are computed using proportional-integral-derivative (PID) control laws

$$U_{as}(t) = K_{PS}\beta_e(t) + K_{DS}\frac{d\beta_e(t)}{dt} + K_{IS}\int_0^t \beta_e(t)dt \quad (3)$$

and

$$U_{ad}(t) = K_{PD}\omega_e(t) + K_{DD}\frac{d\omega_e(t)}{dt} + K_{ID}\int_0^t \omega_e(t)dt, \quad (4)$$

where K_{PS}, K_{DS}, K_{IS}, K_{PD}, K_{DD}, K_{ID} are gains. In addition, $\beta_e(t)$, $\omega_e(t)$ are the errors between the reference variables β_r, ω_r and the feedback measurements β_f, ω_f, given by

$$\beta_e(t) = \beta_r(t) - \beta_f(t) \quad (5)$$

and

$$\omega_e(t) = \omega_r(t) - \omega_f(t). \quad (6)$$
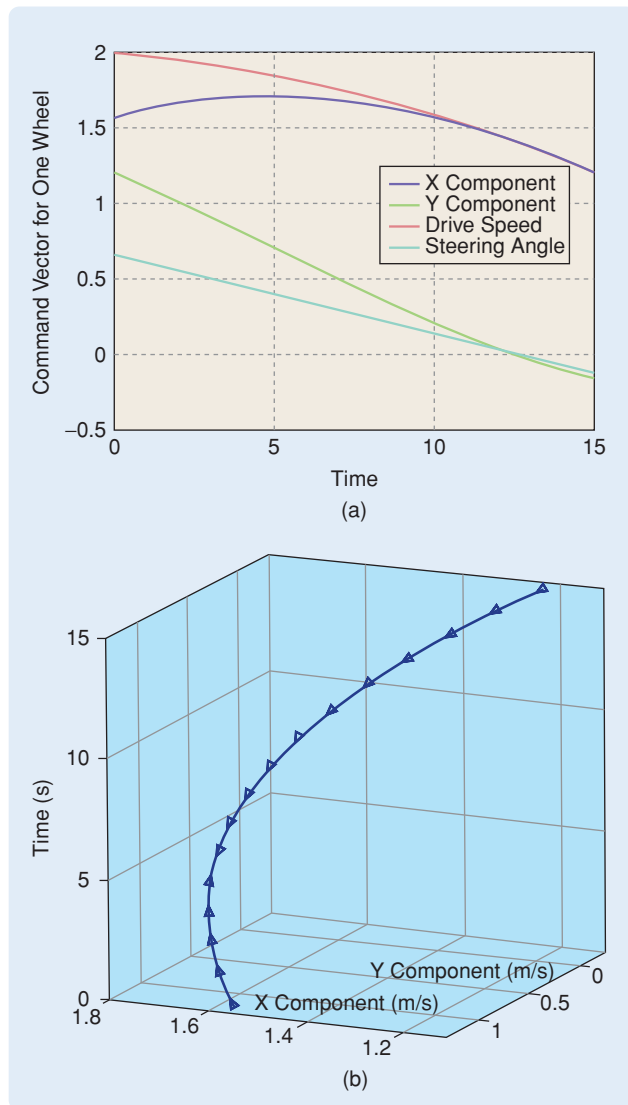
FIGURE 6 Command input value curves. (a) This plot shows the x component, y component, driving speed, and steering angle for the front right wheel of the vehicle. (b) This plot shows the x component and y component of the control input values with respect to time.

Figure 5 shows the attitudes of the wheels and the vehicle at various locations along a reference path. Note that the planar motion of the vehicle consists of translational and rotational motion. While translational motion moves the geometric center precisely along the reference path from A to D, the commanded rotational motion changes the vehicle orientation with respect to the path in a gradual and smooth fashion. From D to E, the planar motion of the vehicle consists only of translational motion since no vehicle rotation is needed.

Figure 6 shows the reference values for the front right wheel from A to D in Figure 5. Both curves of the wheel's driving speed and steering angle are changed gradually with no abrupt variation during this segment. Since the movement is slow, only minimal wheel slippage can occur.

AVOIDANCE OF SLIPPAGE

Synchronization among the four wheels' steering references plays an important role in their coordination. Figure 7 describes command execution state curves for the four steered wheels during a maneuver. PID speed-control algorithms are applied to the four driving motors with parameters that are different from the steer-control algorithms. When all four wheels are steered at the proper angles and precisely driven, the slippage of wheels can be minimized. However, when one wheel does not follow its steer reference command, for example, due to an environmental obstacle, large wheel slippage can occur, and the motion of the vehicle is difficult to estimate. A similar slippage problem occurs when the wheel speed references are not tracked precisely. To solve this problem, we allow the steering and drive controllers to actively report execution states of the steering commands to the upper navigation algorithm through the
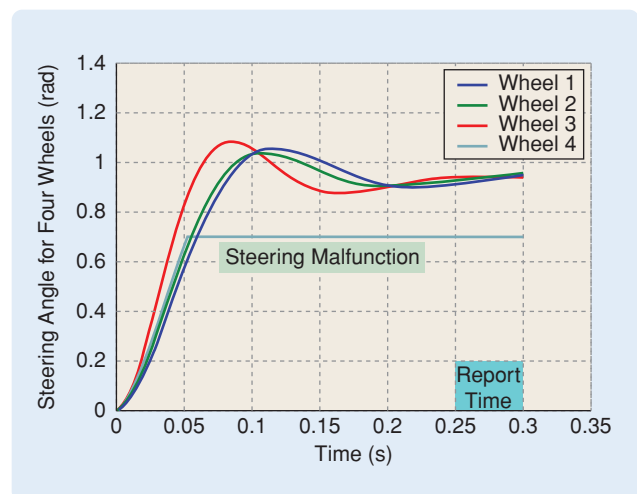
FIGURE 7 Steering command execution states of the four steering motors. The four curves represent the four steering execution states. Although wheels 1, 2, and 3 steer as expected, wheel 4 experiences a malfunction, which limits its steering angle to an angle of 0.7 rad.

vehicle CAN bus. The navigation algorithm waits for the reports. If the steering commands are executed correctly and in a timely manner, the drive commands are sent as usual. If a timeout report is received, the navigation algorithm calls other function modules to handle this situation, during which the information from the IMU, machine vision, and AGDGPS132 are fused.

ACKNOWLEDGMENT

REFERENCES

[1] M. Kise, Q. Zhang, and M.F. Rovira, "A stereovision-based crop row detection method for tractor-automated guidance," *Biosyst. Eng.,* vol. 90, no. 4, pp. 357–367, 2005.
[2] O.C. Barawid, Jr., A. Mizushima, K. Ishii, and N. Noguchi, "Development of an autonomous navigation system using a two-dimensional laser scanner in an orchard application," *Biosyst. Eng.*, vol. 96, no. 2, pp. 139–149, 2007.
[3] J. Borenstein, "Control and kinematic design of multi-degree-of-freedom mobile robots with compliant linkage," *IEEE Trans. Robot. Automat.,* vol. 11, no. 1, pp. 21–35, 1995.

AUTHOR INFORMATION

Chengliang Liu (chlliu@sjtu.edu.cn) received the Ph.D. from Southeast University, NanJing, China, in 1999. He is a professor in the School of Mechanical Engineering at Shanghai Jiaotong University, China. His research interests include agricultural robots, remote monitoring and control, intelligent maintenance, and mechatronics.

Mingjun Wang received the B.Tech. from Shanghai Jiaotong University, China, in 2004. He is currently a Ph.D. candidate in the Department of Mechanical Engineering. His research interests include computer vision, motion control, and robotics.

Jun Zhou received the Ph.D. from NanJing Agricultural University in 2003. He is currently a professor in the department of agricultural machinery in NanJing Agricultural University. His research interests include image processing and computer vision.

Robo Armwrestler

CHUL-GOO KANG

An arm-wrestling robot called Robo Armwrestler (Figure 1) has recently been developed in our Intelligent Control and Robotics Laboratory to benefit the health care of senior citizens. The motivation for this project, which is supported by the Korean Government, is to reduce social welfare costs and to improve the quality of life of the elderly population by meeting their physical and mental needs. In recent decades, Korea's aging population has increased by 35%, well over the standard 30% for an aging society. Our vision is to realize humanoid robots that have entertaining functions such as arm wrestling and chess playing, as well as service functions such as errands. Over time, we hope the robot will help both the elderly and the disabled.

Several years ago, Y. Bar-Cohen issued a challenge to build a robot using muscles of electrically activated polymers that could arm wrestle a human [1]. As a result of this challenge, a few arm-wrestling robots were built using electroactive polymers (EAPs) [2]. The primary object of these arm-wrestling robots is to demonstrate the potential of EAP technology. These robots, however, do not have a broad range of functions related to arm-wrestling skills. Another effort relating to a humanoid robot arm has been in the field of prosthetic devices, such as the Utah Artificial Arm [3]. However, it does not appear that prosthetic devices are suitable for arm wrestling, in which strong arm force is required.

Several practical arm-wrestling devices have been patented as amusement units or as units for developing and strengthening wrist and arm muscles [4]–[7]. These devices are classified roughly into three types according to

FIGURE 1 The arm-wrestling robot Robo Armwrestler. This robot, which is intended for use in elderly health care, was developed in the Intelligent Control and Robotics Laboratory at Konkuk University in Korea.

vehicle CAN bus. The navigation algorithm waits for the reports. If the steering commands are executed correctly and in a timely manner, the drive commands are sent as usual. If a timeout report is received, the navigation algorithm calls other function modules to handle this situation, during which the information from the IMU, machine vision, and AGDGPS132 are fused.

ACKNOWLEDGMENT

REFERENCES

[1] M. Kise, Q. Zhang, and M.F. Rovira, "A stereovision-based crop row detection method for tractor-automated guidance," *Biosyst. Eng.*, vol. 90, no. 4, pp. 357–367, 2005.
[2] O.C. Barawid, Jr., A. Mizushima, K. Ishii, and N. Noguchi, "Development of an autonomous navigation system using a two-dimensional laser scanner in an orchard application," *Biosyst. Eng.*, vol. 96, no. 2, pp. 139–149, 2007.
[3] J. Borenstein, "Control and kinematic design of multi-degree-of-freedom mobile robots with compliant linkage," *IEEE Trans. Robot. Automat.*, vol. 11, no. 1, pp. 21–35, 1995.

AUTHOR INFORMATION

Chengliang Liu (chlliu@sjtu.edu.cn) received the Ph.D. from Southeast University, NanJing, China, in 1999. He is a professor in the School of Mechanical Engineering at Shanghai Jiaotong University, China. His research interests include agricultural robots, remote monitoring and control, intelligent maintenance, and mechatronics.

Mingjun Wang received the B.Tech. from Shanghai Jiaotong University, China, in 2004. He is currently a Ph.D. candidate in the Department of Mechanical Engineering. His research interests include computer vision, motion control, and robotics.

Jun Zhou received the Ph.D. from NanJing Agricultural University in 2003. He is currently a professor in the department of agricultural machinery in NanJing Agricultural University. His research interests include image processing and computer vision.

Robo Armwrestler

CHUL-GOO KANG

An arm-wrestling robot called Robo Armwrestler (Figure 1) has recently been developed in our Intelligent Control and Robotics Laboratory to benefit the health care of senior citizens. The motivation for this project, which is supported by the Korean Government, is

FIGURE 1 The arm-wrestling robot Robo Armwrestler. This robot, which is intended for use in elderly health care, was developed in the Intelligent Control and Robotics Laboratory at Konkuk University in Korea.

to reduce social welfare costs and to improve the quality of life of the elderly population by meeting their physical and mental needs. In recent decades, Korea's aging population has increased by 35%, well over the standard 30% for an aging society. Our vision is to realize humanoid robots that have entertaining functions such as arm wrestling and chess playing, as well as service functions such as errands. Over time, we hope the robot will help both the elderly and the disabled.

Several years ago, Y. Bar-Cohen issued a challenge to build a robot using muscles of electrically activated polymers that could arm wrestle a human [1]. As a result of this challenge, a few arm-wrestling robots were built using electroactive polymers (EAPs) [2]. The primary object of these arm-wrestling robots is to demonstrate the potential of EAP technology. These robots, however, do not have a broad range of functions related to arm-wrestling skills. Another effort relating to a humanoid robot arm has been in the field of prosthetic devices, such as the Utah Artificial Arm [3]. However, it does not appear that prosthetic devices are suitable for arm wrestling, in which strong arm force is required.

Several practical arm-wrestling devices have been patented as amusement units or as units for developing and strengthening wrist and arm muscles [4]–[7]. These devices are classified roughly into three types according to

the methods of generating reaction force against a player's arm force.

The first type of arm wrestling device uses a spring force; a typical example is U.S. Patent 3,947,025 [4], in which the device is comprised of a helical coiled spring that has adjustable stiffness. The second type of device uses pneumatic or hydraulic cylinders, which is better than the spring type from the viewpoint of force manipulability. However, a disadvantage is that the system becomes complicated and bulky because of the need for a supplementary device for pneumatic or hydraulic pressure generation. An example of the hydraulic device is U.S. Patent 5,842,958 [5]. The third type of device uses electric motors instead of springs or pneumatic/hydraulic cylinders in order to generate resistive force against the user. Examples include Japan patent 6-315544 [6] and Japan patent 2002017891 [7], in which torque motors are used to generate arm force, and sensor plates and photosensors are used to detect arm speed to prevent a throw fracture of the player.

These patented devices are invented for playing simple arm-wrestling games or for practicing strength training, in which they usually generate fixed force levels. If the player generates a stronger force than the arm wrestling device, then he will win; otherwise he will lose. Unfortunately, players become easily bored after a few trials.

Robo Armwrestler has features that remedy the above deficiencies. In particular, Robo Armwrestler automatically generates the force level appropriate to each person after sensing the human arm force at the early stage of the match, and therefore any person with either a strong or weak arm force can enjoy arm wrestling without any parameter changes. Furthermore, Robo Armwrestler's generated force profile varies with each match, so each person can enjoy arm wrestling with the robot for a long time without becoming bored. The winning average of the robot is determined randomly at the start of the match, but the robot measures the person's will to win during the match, and this result influences the winning average of the robot. For example, the robot's arm force becomes weaker and the winning probability of the human increases if the human tries to win the match over a period of time.

The robot recognizes a human's approach and begins to talk, saying, for instance, "Hello, welcome to Robo Armwrestler! If you want to try arm wrestling, please sit down on the chair." The device also recognizes the person sitting on the chair and guides them on how to play and enjoy arm wrestling. The tone of the voice varies occasionally so as not to be tedious. The facial expression of the avatar changes synchronously according to arm-wrestling situations.

HARDWARE DESIGN

Robo Armwrestler can detect the maximum arm force of the user in the early stage of a match and can automatically and randomly generate a different arm-wrestling sce-

nario each time so that the user cannot predict a force pattern in advance. The robot then executes force feedback control to implement the scenario using signals related to the motion of the mechanical arm and to the torque acting on the mechanical arm. The arm-wrestling robot is thus composed of an arm-force generation mechanism, a control system, and an inference engine.

The arm-force generation mechanism comprises a servomotor, a position/velocity sensor, a speed reducer, a torque sensor, three inclinometers, a mechanical arm, and an adapter with a mechanical stopper [8]. The electric servomotor provides necessary torque according to motor control input signals, while the position/velocity sensor detects angular position and angular velocity of the motor. The speed reducer decreases the speed and increases the torque of the motor. The torque sensor detects the torque acting on the mechanical arm, and the adapter with a mechanical stopper is utilized to restrict the angular range of motion of the mechanical arm to guarantee the safety of the user. The inclinometers are used to initialize the absolute angle of the robot arm. The arm-force-generation mechanism makes the mechanical arm rotate in either a clockwise or a counterclockwise direction.

An incremental encoder is used as the position/velocity sensor for high resolution. A harmonic drive is selected as the speed reducer instead of conventional gears, which have large backlash and thus limit torque control performance. The adaptor with the mechanical stopper is further utilized to set the initial absolute angle of the mechanical arm by means of low-speed control of the motor. The initial setting of the absolute angle of the arm can also be achieved redundantly by using multiple inclinometers.

The torque sensor installed between the speed reducer and the mechanical arm must have reasonable resolution to obtain reasonable force-control performance. Two ultrasonic sensors, which are attached at the right and left sides on the front of the table, detect a human's approach within a prescribed range of angles near the arm-wrestling robot. Ultrasonic sensors have the advantage of high noise immunity compared to other types of sensors and can easily measure the distance of an approaching human under any circumstances. One photoelectric sensor using infrared rays detects the human sitting on a chair. To guide the player, the display monitor and speakers are visibly positioned on the table.

The arm-force generation mechanism is shown in Figure 2, while the schematic of Robo Armwrestler is shown in Figure 3.

FORCE CONTROL SCHEME

To make the arm-wrestling game attractive, we add force control functions to the system in such a way that force command is generated intelligently and so that real-time tracking control can follow the given force command. Linux and RTAI (real-time kernel) are adopted for the operating system of the robot. When Linux and RTAI are

implemented on a desktop PC with a 2.8-GHz Pentium IV, timing errors less than ±0.02 ms occur in generating 5-ms timer interrupts. However, Windows XP has roughly ±1-ms latency when the same program is executed with the same PC platform.

A block diagram of the force feedback control scheme is shown in Figure 4. The control system receives feedback signals of actual position, velocity, and torque and calculates torque command using feedback signals and scenarios. The system then controls the mechanical arm by generating the motor control input using force control logic. The force control logic is basically PID type but uses velocity and position information together with force information.

Force control performance depends mainly on the accuracy of the feedback signals and real-time control capability, including the accuracy of the sampling time, as well as the force feedback control logic itself. Force feedback control plays a key role in arm wrestling the robot, but position feedback control is also necessary for rotating the mechanical arm to a starting position and setting the initial absolute angle of the mechanical arm before a match. For this servo control, the motor driver supplies analog voltage signals corresponding to angular speed as well as quadrature encoder signals (A, B, and Z). We generate digital position and velocity signals using the quadrature signals, since the generated digital signals are more accurate than the analog voltage signals.

As shown in Figure 5, the control system of the robot is comprised of a CPU, memory, amplifier, logic circuit, pulse-generation circuit, and output ports. The CPU produces a motor control input using the control program and the feedback signals and produces voice and image signals. The memory stores a control program including control logic and scenarios. The amplifier amplifies the low-level voltage signal coming from the torque sensor and achieves signal conditioning. The logic circuit conditions the feedback signal from the position/velocity sensor, while the pulse-generation circuit produces a pulse signal for the ultrasonic sensors. Voice speakers and a display monitor are driven by the CPU through output ports.

The torque sensor, inclinometer, photoelectric sensor, and ultrasonic sensor signals are converted to digital signals through analog-to-digital (A/D) converters and transmitted to the CPU. Encoder signals are transmitted to the CPU directly through digital input channels. The motor driving signals are converted to analog voltages through a digital-to-analog (D/A) converter and transmitted to the motor driver.

When the CPU is down, the D/A converter can still output the last signal of the motor control input, and thus a dangerous situation can occur if electric power is applied to the motor. To resolve this problem, the

FIGURE 2 Mechanism for arm-force generation. The mechanism is composed of a servomotor, a position/velocity sensor, a torque sensor, three inclinometers, a speed reducer, a mechanical arm, and an adapter with a mechanical stop.

FIGURE 3 Perspective view of Robo Armwrestler. This view includes an arm-force generation mechanism, a bust, a table and chair, two ultrasonic sensors, and a photoelectric sensor.

FIGURE 4 Simplified block diagram of the control system. The control system generates a torque command at each sampling time using position, velocity, and torque information, and then accomplishes force feedback control using PID-type logic.

CPU transmits an initialization completion signal to the motor power control through a D/A converter and sends a zero value to the motor drive through the D/A converter when the initialization procedure is completed, where the initialization procedure starts when the main switch is pressed. The motor power controller turns on the mechanical relay (MR) to supply electric power to the motor according to the output signal of the solid-state relay (SSR), which in turn is actuated by the initialization completion signal.

User safety is thus guaranteed even if the motor power switch is turned on before completing the initialization procedure or at abnormal conditions of the control system since electric power is not transmitted to the motor.

FIGURE 5 Hardware configuration of the control system. Torque sensor, inclinometer, photoelectric sensor, and ultrasonic sensor signals are transmitted to the CPU through A/D converters. Encoder signals are transmitted through digital input channels. The solid-state and mechanical relay circuits in the motor drive guarantee user safety.

When using the incremental encoder as the position/velocity sensor, we initially set the absolute zero-degree angle of the mechanical arm using the mechanical stopper and velocity feedback control. More specifically, the controller slowly drives the motor clockwise or counterclockwise using position feedback control and measures the torque value by means of the torque sensor. If the measured torque is larger than the threshold value, then the controller sets the present angular position as the absolute angle, since the large measured torque implies that the mechanical stopper has reached the stopper seat block.

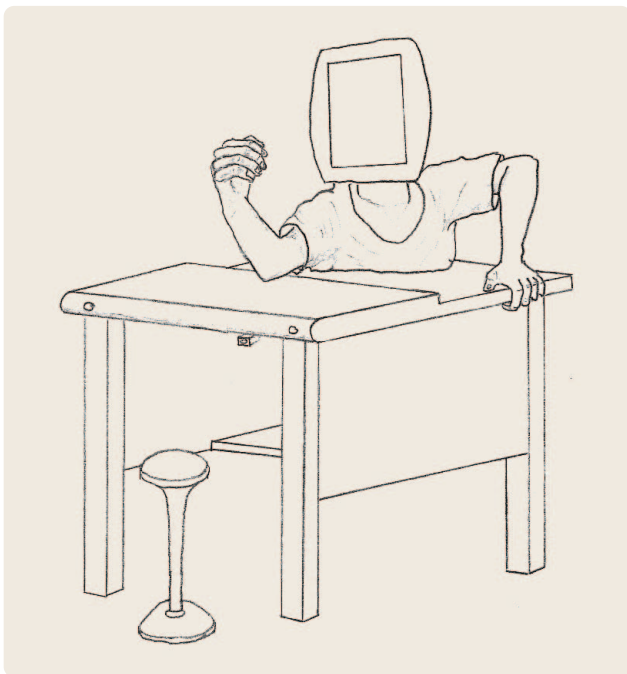Initial setting of the absolute arm angle can also be accomplished redundantly using multiple inclinometers, but in this case the arm-force generation mechanism becomes more complicated, and possibly more expensive.

INTELLIGENT SCENARIOS

Arm wrestling is not strictly a strength sport, as people often think, because technique and speed are also very important. Therefore, a key element of an arm-wrestling robot is to create an intelligent game scenario online. In Robo Armwrestler, the inference engine generates an appropriate force profile for each match, which considers a human's maximum force, force pattern, time duration, will to win, and randomness. Force control logic enables the robot arm to follow the generated force profile as smoothly as possible. In Robo Armwrestler, winning or losing is not predetermined but varies online according to the pattern of the game progression.
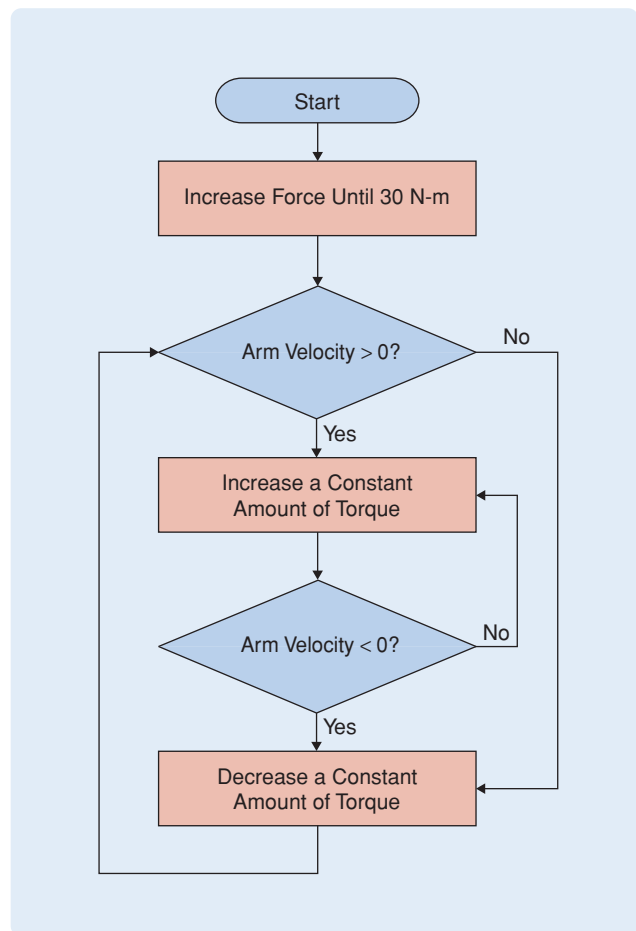
FIGURE 6 Procedure for measuring maximum human force. This procedure is executed for a few seconds in the early stage of each match.
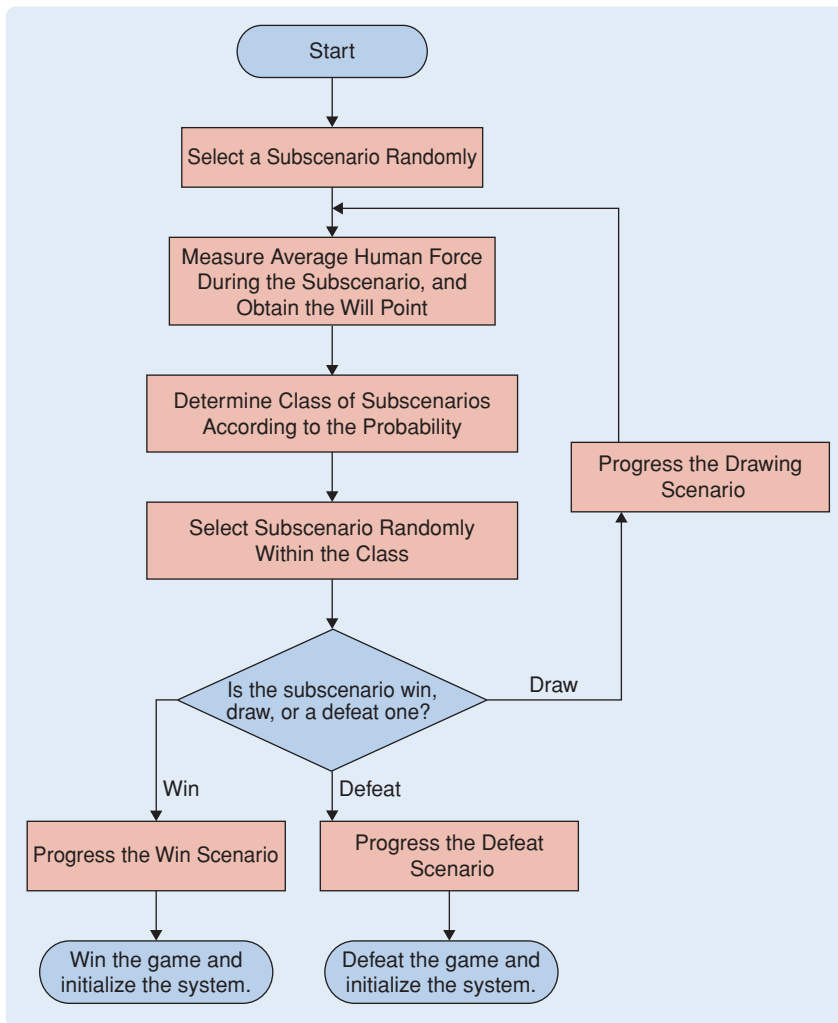
FIGURE 7 Flowchart showing the progression of robotic arm wrestling. The winning average is determined by random numbers and will points.

For a few seconds in the early stage of the game, the inference engine detects the human's maximum force according to the procedure shown in Figure 6. The system increases force up to the specified value with a parabolic fashion for a short time, and then the robot measures arm velocity every 0.1 s for the next few seconds. If the velocity is positive, then the robot increases the force until the velocity becomes negative, and records the force value.

To realize unpredictable and intelligent game patterns, we adopt a random number and a value called *will point*, which quantifies the determination of the arm wrestler to win the match. If the will point is near 100, the user is considered to have a strong desire to win the match. If the will point is near zero, the user is considered to have a weak desire to win the match. The will point is calculated by

will point =(average arm force during one subscenario)/

(maximum arm force of the user) × 100.

The game scenario is composed of several subscenarios. Each subscenario has a different rotation angle limit of

the robot arm within 150° and is divided into three classes: win, draw, and defeat. If a subscenario is selected by means of random number generation, the robot decreases or increases the force during randomly determined intervals of time. During these intervals, human force is measured and averaged to calculate the will point. As soon as the execution of a subscenario is completed, the next subscenario is immediately prepared. This subscenario can be generated online at that instant or can be selected among many subscenarios prepared in advance.

The win subscenario implies a significant decrease of the torque command value, the defeat subscenario implies a significant increase of torque command value, and the draw subscenario implies a small increase, a small decrease, or no change of torque command value. The grouping of win, draw, and defeat subscenarios depends on the present arm angle, and is achieved using a prescribed rule.

Arm-wrestling progression is affected by the will point and pre-specified probability. For example, if a will point of 86 is obtained, then the class of win, draw, or defeat scenario is determined according to winning probability with 8%, drawing probability with 90%, and defeat probability with 2%. This class determination is conducted using a random number between 0 and 99, that is, the generated random numbers 98 and 99 imply the defeat class, random numbers 8 to 97 imply the drawing class, and random numbers 0–7 imply the winning class.

Using another random number, we select a subscenario randomly within subscenarios of the determined class. If the selected subscenario is a drawing one, then the will point is recalculated after the subscenario ends, and the above procedure is repeated. If the selected subscenario is a win or a defeat, then the win or defeat subscenario is employed, and the human wins or defeats, and the match ends. Finally, the arm-wrestling system is initialized for the next match. Figure 7 shows the flowchart of these arm-wrestling scenarios.

Instead of selecting a subscenario among three classes, we can generate a subscenario online, which is characterized by three parameters, force increment, rise time, and maintenance time, as shown in Figure 8. These parameters are determined using a random
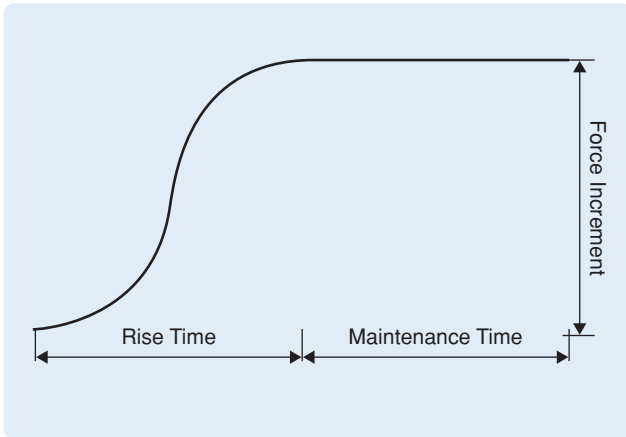
FIGURE 8 Force increment pattern. The pattern is composed of a smoothly increasing force interval and an interval of constant force.

(a)

(b)

(c)
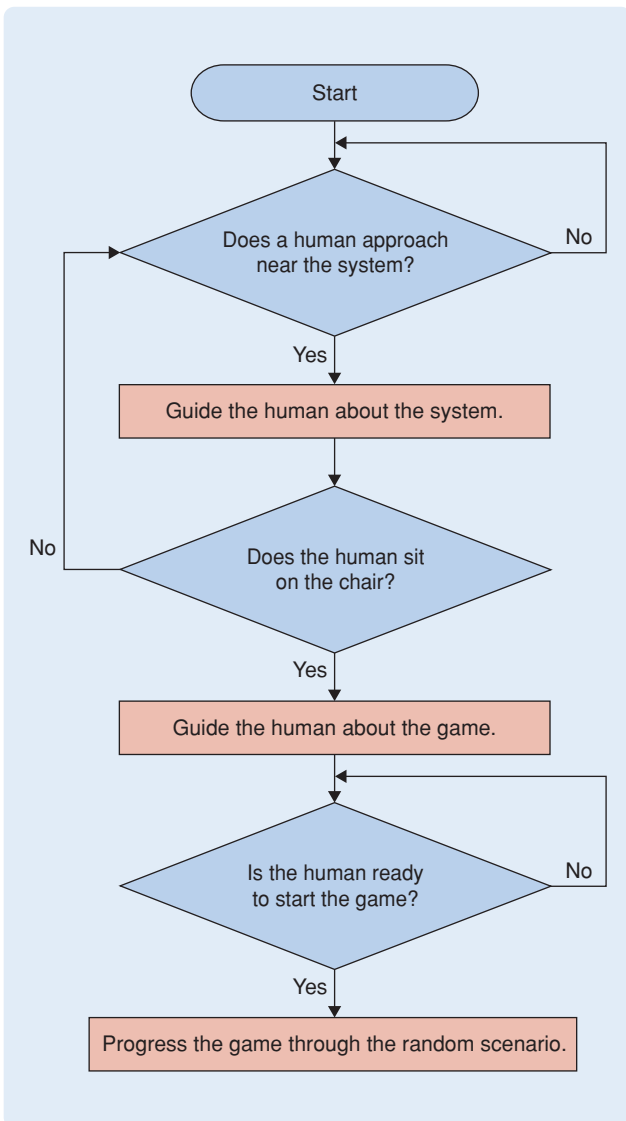
FIGURE 9 Beginning procedure of robotic arm wrestling. This procedure includes automatic recognition of a human's approach to the robot, as well as the human sitting in the chair.
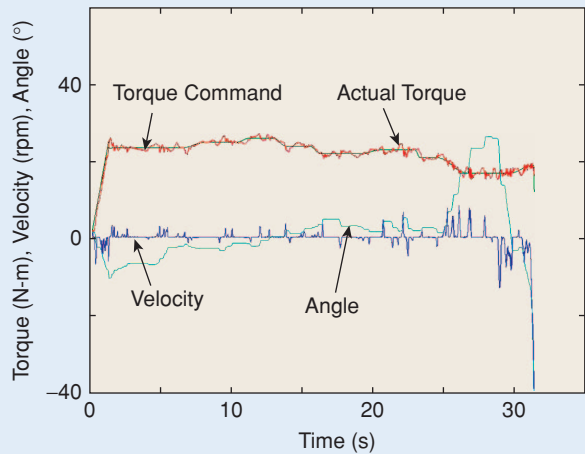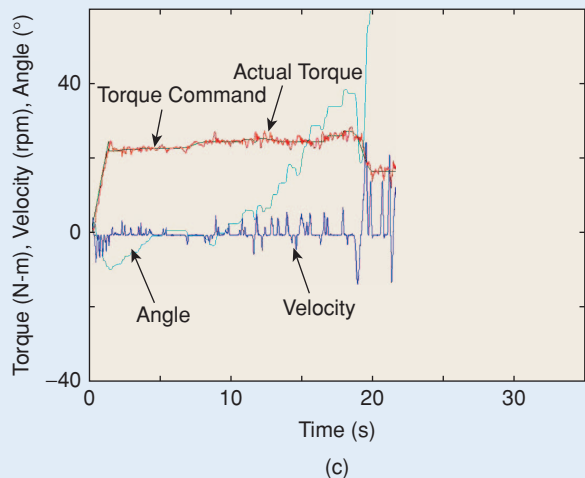
FIGURE 10 Robotic arm-wrestling results of a 72-year-old woman. (a) Arm-wrestling picture, (b) a case where the woman wins the game after 31 s, (c) a case where the woman loses the game after 21 s.
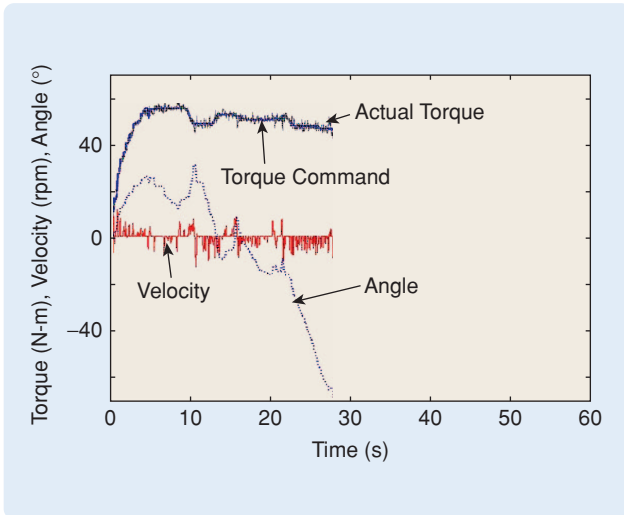
FIGURE 11 Outcome of a young adult arm wrestling. In this robotic arm-wrestling contest, the 25-year-old adult loses the game after 28 s.

(a)

(b)
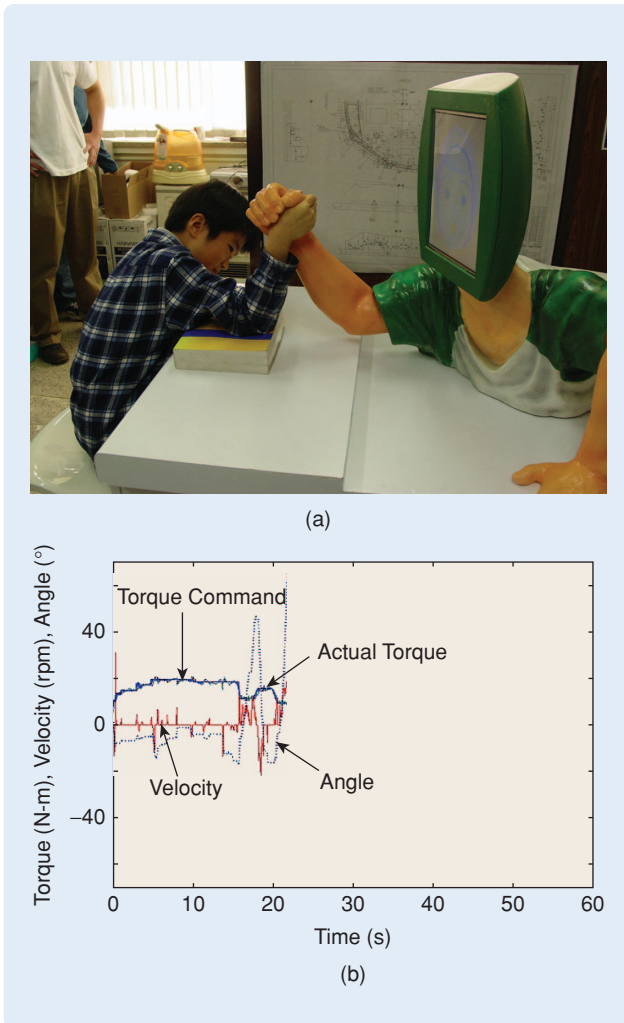
FIGURE 12 Robotic arm-wrestling result of a ten-year-old boy. (a) Arm-wrestling picture and (b) a case where the boy wins the game after 23 s.

number and the will point. The force increasing or decreasing in the subscenario is achieved by smooth polynomial curves as shown in Figure 8.

During an idle situation, the robot plays music and waits for a person to appear. As a person approaches the robot, the system automatically detects the approach, greets him or her, and encourages the person to arm wrestle. If the person sits down, the robot guides him/her to start the game. This procedure is shown in Figure 9.

DEMONSTRATION

Experiences at the laboratory and the Future Tech Korea Exhibition reveal that Robo Armwrestler generates intelligent scenarios unpredictably and reliably and controls the robot arm force properly so that the human arm wrestler feels as if he or she is arm-wrestling against a human.

Figure 10(a) shows a 72-year-old woman playing against Robo Armwrestler. Figure 10(b) and (c) shows match results of Figure 10(a), in which the torque command, angular velocity, and arm angles are plotted with respect to time. Figure 10(b) corresponds to the case in which the woman loses, while (c) corresponds to the case in which the woman wins. From these graphs, we can see that force patterns generated by the arm-wrestling robot and the elapsed time of arm wrestling are different from match to match even if the same person plays.

Figure 11 shows a result of the match between a strong 25-year-old adult and Robo Armwrestler. Figure 12(a) and (b) illustrates a match between a ten-year-old child and Robo Armwrestler. As soon as the adult finishes the match in Figure 11, the child begins the match shown in Figure 12(a). Although the adult produces roughly 50 N-m and the child roughly 20 N-m, the arm wrestling proceeds smoothly without changing any parameters of Robo Armwrestler. In other words, Robo Armwrestler generates an appropriate force level automatically according to the magnitude of each player's arm-force.

Table 1 shows the elapsed time and winning/losing probability of each match when one user plays arm wrestling ten times with Robo Armwrestler. We can see that the elapsed time for a match varies each time, and the result of the match also varies each time, so that the enjoyment of arm wrestling can be maintained for a long time.

TABLE 1 Game statistics. These elapsed times and match results correspond to one user playing ten times with Robo Armwrestler.

Trials	Elapsed Time	Result	Trials	Elapsed Time	Result
1	11.0 s	Win	6	17.6 s	Lose
2	14.0 s	Win	7	14.5 s	Lose
3	20.0 s	Win	8	7.8 s	Lose
4	47.0 s	Lose	9	23.4 s	Win
5	20.0 s	Win	10	11.5 s	Win

Table 2 summarizes results for two users arm wrestling 26 times each with Robo Armwrestler. The table shows that the first user has a 63% probability of winning, while the second user has a 75% winning probability.

CONCLUSIONS

In this article we have presented the system design, implementation, force feedback control, and generation of intelligent arm-wrestling scenarios of an arm-wrestling robot. Although Robo Armwrestler works as expected with the designed autonomy and reasonable control performance, we plan to further pursue research to provide the device with the ability to recognize facial expressions of a human using a Webcam and thus emotionally communicate with the player. Moreover, we plan to add degrees-of-freedom for more human-like motion and to eventually integrate arm-wrestling functions into a humanoid robot.

ACKNOWLEDGMENTS

REFERENCES

[1] Y. Bar-Cohen, "Electric flex," *IEEE Spectr.*, vol. 41, no. 6, pp. 28–33, 2004.
[2] G. Kovacs and P. Lochmatter, "Arm wrestling robot driven by dielectric elastomer actuators," *Proc. SPIE*, vol. 6168, pp. 1–12, 2006.
[3] E. Iversen, H.H. Sears, and S.C. Jacobsen, "Artificial arms evolve from robots, or vice versa?" *IEEE Control Syst. Mag.*, vol. 25, no. 1, pp. 16–20, 2005.
[4] J.M. Hobby Jr., "Arm wrestling unit," U.S. Patent 3947025, 2004.
[5] F.P. Rufa, "Arm wrestling device," U.S. Patent 5842958, 1998.
[6] T. Giken and K.K. Kogyo , "Hand wrestling game machine," Japan Patent 6315544, 1994.
[7] S. Toshio, "Muscular force strengthening device for arm wrestling," Japan Patent JP2002017891, 2002.
[8] C.-G. Kang and H.-Y. Kim, "Motion-dependent force control for an arm-wrestling robot," in *Proc. 13th Int. Conf. Advanced Robotics*, Jeju, Korea, 2007, pp. 999–1004.

	Elapsed Time Average	Percentage of Human Wins
Person 1	17.0 s	63%
Person 2	14.0 s	75%

TABLE 2 Two-person statistics. These results correspond to two users playing 26 times each with Robo Armwrestler.

AUTHOR INFORMATION

Chul-Goo Kang (cgkang@konkuk.ac.kr) received the B.S. and M.S. in mechanical design and production engineering from Seoul National University, Korea, in 1981 and 1985, respectively, and the Ph.D. in mechanical engineering from the University of California, Berkeley, in 1989. He joined Konkuk University, Seoul, Korea, in 1990, where he is a professor and the director of the Future Robot Research Center and the Intelligent Control and Robotics Laboratory. His research includes intelligent robots, force/torque sensing, and brake systems of high-speed trains. He is an editor of the *Journal of the Korea Fluid Power Systems Society*, a board member of the Korea Robotics Society, and a Congress Highlight program editor of the IFAC 2008 World Congress.

Global Challenge

The hunting problem was grave, of course, because it threatened the productivity (quantitative and qualitative) of nations. Tens of thousands of engines worldwide were misbehaving, and so the mathematicians, scientists, and engineers of several countries sought a solution to the problem. One promising approach was to design an improved governor that eliminated hunting entirely, instead of simply trying to mitigate the bad behavior, by eliminating the source of hunting. The source, ultimately, was the way in which engine speed depended on load. If a steam engine/governor system could be designed so that engine speed was independent of load, then hunting would not arise. Much effort went into designing such engine/governor systems. The German William Siemens and his brother spent over 30 years designing governors, trying to solve the problem of hunting. He described a related problem associated with the Watt type of governor, that of offset. He said that the Watt governor "... cannot regulate, but only moderates the velocity of the engine, that is, it cannot prevent a permanent change in the velocity of the engine when a permanent change is made to the load upon the engine" (Bennett). An astatic governor, one that yielded constant engine speed whatever the load, would solve this problem as well as hunting. The French physicist M. L. Foucault devoted much thought to the design of astatic governors.
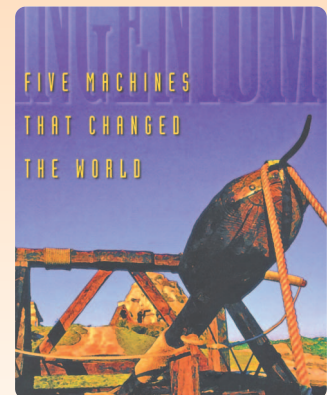
—M. Denny, *Ingenium: Five Machines That Changed the World*, Johns Hopkins University Press, Baltimore, MD, 2007, pp. 146–147.

In this issue of *IEEE Control Systems Magazine* (*CSM*), we speak with Elena Zattoni, assistant professor in the School of Engineering at the University of Bologna, Italy. Elena has worked on a wide range of linear and nonlinear problems for applications that include automotive and railway systems. She is currently the standing chair for the IEEE Control Systems Society Standing Committee on Women in Control.

We also speak with Eyad Abed, who is a professor in the Department of Electrical and Computer Engineering at the University of Maryland, College Park. Abed is serving a two-year term as vice president for financial activities of the IEEE Control Systems Society (CSS). He has also been director of the Institute for Systems Research at the University of Maryland since 2002.

ELENA ZATTONI

Q. **One of your interests is fault detection. What motivated your interest in this problem?**

Elena: A great deal of my research activity has been focused on basic problems of control theory, such as reference tracking, rejection of various classes of signals, output regulation for nonminimum-phase systems, and the dual problems of observation and filtering. Hence, when I first approached fault detection, it appeared to me as an intriguing theoretical problem, an extension of the left-inversion and unknown input observation problems.

Almost contemporaneously, I reviewed an invited session on structural health monitoring for a major control conference. That work, although carried out in the field of civil engineering, helped me to understand how crucial it is to develop valuable tools for health monitoring of infrastructures to guarantee higher levels of safety for individuals, respect of the environment, and, at the same time, affordable maintenance costs, the latter making it all really feasible.

When my research group and I received the proposal to collaborate with an industrial partner to study innovative systems for health monitoring of railway signaling devices, I

was already aware of the relevance of the issue and ready to accept the challenge. For a long time since then, the case study attracted most of our interest and became the benchmark for testing our methodologies and the motivation for developing new ones.

More specifically, the objective of our studies was the synthesis of computational algorithms for detecting faults of an incremental nature in electric switch machines for railway points, namely, faults connected to progressive increasing of frictional loads due to loss of lubrication, deterioration of slide chairs, and increasing obstructions. We developed several solutions, some relying on extensive use of structural and geometric tools,

others essentially based on optimization techniques.

In particular, the main detection system we proposed was based on the offline processing of armature current and voltage of a permanent-magnet dc motor, sampled as the machine operated a switch for railway tracks. Those data are normally available from conventional railway signaling infrastructures, therefore, no additional transducers or instrumentation were required. The algorithm, which was tuned on a linearized model of the machine behavior as both the rails were simultaneously driven toward their respective rest positions, realized a set of finite impulse response (FIR) systems. The FIR system convolution profiles were designed on the basis of an H_2-norm criterion that guarantees robustness, particularly with respect to electrical noise.

Although our work shared the same offline option with recent literature on railway signaling infrastructures, it differed in the methodology adopted to solve the problem. In fact, according to a policy in the geometric framework, the design of the residual generators was performed in the dual context of noninteraction so that the

Elena Zattoni, assistant professor in the School of Engineering at the University of Bologna, Italy, walking on Piazza della Mercanzia in Bologna. On the left is a partial view of the front doors and wooden porch of Seracchioli House (seventh century).

main constraints introduced by practical implementation requirements could be taken into account in terms of finite control horizon, finite control energy, and zero final state.

Q. How are the faults detected?

Elena: In general, structural and geometric approaches to control and observation theory are inherently centered on the mathematical form of systems. Hence, they are enlightening particularly in the early stages of the synthesis procedure, namely, in analyzing the constraints that the intrinsic structure of systems imposes on achieving certain targets as well as in outlining the preliminary design of regulators and observers. Indeed, geometric and structural approaches tend to provide sets of necessary and sufficient conditions for solvability of the specific problem at issue, such as rejection and right inversion, in exact terms. Those conditions are formulated in a sharp fashion and are straightforward to check by means of computational tools, given the mathematical description of the systems involved. This formulation allows us to determine the best achievable performance, given the system description, at least in ideal conditions, for example, in the absence of uncertainties, unknown disturbances, or time delays. If the above-mentioned exact conditions are satisfied, geometric and structural approaches directly provide the tools for the synthesis, for instance, the *if*-part of the proofs of many well-known theorems on decoupling or regulation is constructive.

As a matter of course, at a later stage of the design, one should analyze in depth how nonidealities affect the idealized results. Nonetheless, geometric and structural approaches provide powerful means for dealing with specific circumstances. For instance, as to switched linear parameter varying systems, the maximal robust controlled invariant subspace defined for that set of switched systems is the key to feedforward units and feedback regulators that guarantee zero or minimal regulation transients. Another paradigmatic example is represented by time-delay systems. In fact, specific procedures allow us to transform time-delay systems into systems whose coefficients are defined over rings of polynomials. Therefore, structural and geometric results can be applied on those equivalent systems over rings.

As to fault detection in particular, given a mathematical description of the system to monitor, geometric conditions immediately allow us to check whether standard, infinite-horizon fault detection is achievable or not. If those conditions are not satisfied, milder conditions related to the specific indices of the system subblocks allow us to check whether at least finite-horizon fault-detection can be obtained or not. Alternative structural conditions relate partial fault detection to the system Markov parameters. If the milder conditions are not met, then we can use other techniques, such as optimization.

Q. You've also done a lot of work on preview control. What is preview control, and what are its applications?

Elena: Preview control consists of a broad collection of methodologies aimed at solving tracking or rejection problems where the signals to be tracked or rejected are known a priori by a certain amount of time. I refer to tracking and rejection almost similarly since, as is well known, tracking and rejection problems are equivalent by means of an appropriate redefinition of the to-be-controlled variables. Preview control problems can be stated either in the exact context or in the optimal context. In the former case, the control target guarantees that external signals be either perfectly tracked by outputs or totally decoupled from outputs. In the latter case, the aim of control is to achieve the minimal tracking error or the minimal effect of the external signals on the outputs according to a suitably cho-sen, criterion, typically, a norm criterion. The exact formulation and the optimal formulation of preview control problems have both received significant attention from the control community since the early 1970s.

In particular, we can focus on the problem of achieving an internally stable right inverse of a dynamical system in the presence of unstable zeros. It is well known that the problem has intrinsic limitations that can be overcome only by suitably exploiting additional information on the signals to be tracked, namely, preview. The problem formulated in the exact context allows us to make a subtle distinction between finite preview and infinite preview, where the former is basically related to the system relative degree, or to the properties of a suitably defined, conditioned invariant subspace, and where the latter is related to the presence of unstable zeros. Indeed, in a sharper analysis, it is found that the latter is related to the presence of unassignable unstable internal eigenvalues of a specific subspace, the minimal self-bounded controlled invariant subspace.

As a matter of fact, feedforward units exploiting the infinite preview of the signals to be tracked, although well-defined entities from a theoretical point of view, are not practically implementable since the unavoidable truncation of convolution profiles generates tracking errors. Hence, many studies have been carried out on this issue. Alternative optimal approaches leave the above-mentioned distinctions apart from consideration and provide feedforward compensators that guarantee minimal error, according to the cost functional that has been chosen, given the available preview of the signals to be tracked.

Applications of preview control are manifold, ranging from flight control to manufacturing control and also to signal processing in the dual counterpart represented by smoothing. Recently, I have considered the extension of preview control to the synthesis of

controllers that guarantee output regulation in the presence of swift parameter changes of the to-be-controlled systems, on the assumption that the switching law be known with a given preview. There are many circumstances where the variations in the system dynamics and the time of their occurrence are known in advance. For instance, in aircraft flight control, some maneuvers are predetermined, which implies that switching between the linear time-invariant system modeling the aircraft dynamics at different points of interest throughout the operational envelope can be specified a priori. In chemical processes, the dynamics of certain reactions, under different laboratory conditions, can be predicted. In control of machine tools, the profiles to be tracked, which are normally available ahead of their processing, can be obtained through a set of switched exosystems.

Q. **Are you exploring any new research areas?**

Elena: My current research deals with the investigation of geometric methods for output regulation in systems subject to swift parameter variations, known with given preview as to amplitude and time of occurrence in the more general case where the to-be-controlled systems may include time delays.

The study of this issue requires competence in a variety of topics, ranging from preview control to linear parameter varying systems and time-delay systems. Hence, it benefits from the collaboration of multiple research groups that share the geometric approach as a common, privileged methodology. In fact, the geometric approach has proved to be effective in coping with systems subject to parameter jumps as well as in treating time-delay systems.

As to time-delay systems in particular, the possibility of applying geometric methods based on a procedure is well documented in the control literature, which associates an abstract system with coefficients in a suitably defined ring of polynomials to any time-delay system. Consequently, control problems originally addressing time-delay systems implying infinite-dimensional state-space representations can be recast in an algebraic context where the system behaviors are represented by standard, discrete-time linear systems involving finite-dimensional state space realizations. This way of operating has allowed us to prove sufficient conditions for the solution of the multivariable autonomous regulator problem when the to-be-controlled system is affected by time delays, and it has also led us to the formulation of a complete synthesis procedure.

The extension of results available for systems with coefficients in the real field to systems whose coefficients are defined over rings of polynomials appears relevant and urgent, now that time-delay systems, always present in industrial applications involving transportation effects, are spreading in more recent applications such as networked control systems, teleoperated systems, and integrated communication control systems.

Q. **What courses do you teach at the University of Bologna? Do you have any special teaching techniques that you wish to share?**

Elena: I am currently a professor of control systems within a B.Sc. in electronics and telecommunications engineering, and I am coresponsible for an introductory course on geometric approach to system and control theory for Ph.D. students in automation and operations research.

As to the second part of your question, I should premise that it is my belief that competent and clear explanations from the lecturer and good reference books are the main elements of an effective course at the university level. Indeed, I think that the lecturer should outline the essentials of the subject on the one hand and provide the students with interesting hints and references to deepen on the other. I pay a lot of attention to the organization of my lessons and to the selection of the textbooks. As a standard teaching tool, I also accompany the development of the course with my collection of exercises, specifically tuned on the final exam.

In addition, I firmly believe that a special teaching technique consists of laboratory exercises. In fact, I encourage my students to master a CACSD software specifically focused on the synthesis of single-input, single-output feedback control systems by means of transfer functions. The specific software is named Transfer Function Interpreter (TFI). The students are provided with a collection of ten exercises on the main topics of the course, such as Laplace transforms, root loci, Bode, Nyquist, and Nichols diagrams, phase lead and phase lag design, and PID controllers.

Certainly, laboratory activities are quite demanding, especially with numerous classes, as is often the case with basic courses. However, in favorable conditions I have obtained very good results from most of the students. This is even more true with regard to the Ph.D. course on the geometric approach. In fact, a collection of Matlab functions for computing the resolving subspaces of the main problems and a set of synthesis procedures are of great help to students in becoming more familiar with such an (allegedly) unfriendly methodology.

Q. **What motivated your interest in control as an undergraduate?**

Elena: It is my belief that the general scope of automatic control as an engineering discipline, namely, the study and definition of methodologies and techniques aimed at replacing human resources in performing repetitive tasks as well as in improving human performance and safety in carrying out complex or dangerous activities, is noble in principle, in the sense that all of society can benefit in terms of quality of life from new achievements in this field. Incidentally,

avoiding unvaried occupations and taking measures to assist myself in difficult or risky actions have always been priorities in my own lifestyle. Certainly, some of those ideas played a role when I selected courses on automatic control from the wide-spectrum of offerings of the studies in electronics engineering of the early nineties.

Another attractive feature of automatic control is being an interdisciplinary science. As we all know, principles and technologies of this science are present in a wide variety of applications in different areas. The need for mastering concepts coming from other domains, which, therefore, is a prime concern of the project engineer dealing with control systems, is indeed intriguing and challenging at the same time.

Interaction with different subjects, mechanics in particular, was an essential element of my M.Sc. thesis. In fact, my thesis concerned the study of narrow-bandwidth active suspension systems for high performance conventional cars. Naturally, the study of the hydraulic system, which represents the to-be-controlled system, absorbed a relevant part of my work. I should also admit that world-renowned brands of motor constructors, past and present, have flourished in the region where I live, including Ferrari, Maserati, Bugatti, Lamborghini, and Benelli. Interacting with matters of interest to some of these companies definitely created a mix of tradition and novelty that was fascinating to me.

Q. **You're quite active in the CSS as chair of the Standing Committee on Women in Control. What are some activities of this group?**

Elena: According to its statement, "the Women in Control Committee promotes membership, gathers and disseminates appropriate information about women in the IEEE Control Systems Society and the profession and facilitates the development of mentoring and programs to promote

the retention, recruitment, and growth of women CSS members."

Although women are now involved in many aspects of the CSS, less inclusive environments are still a reality in many workplaces all over the world. Hence, there are still reasons for asking ourselves what actions should we focus on, from the inside of our Society, to guarantee to women scientists and engineers the opportunity to achieve the career attainments they deserve, in terms of research performance and status.

The first answer can be found in the extensive and effective use of the basic means we all have access to, namely, the use of e-mail and the Internet, to create and circulate occasions for women to join research groups, to collaborate, and to gain access to human and material resources. As we all know, the dissemination of this kind of information is crucial in scientific and engineering work.

The fastest and most effective means for spreading information is our mailing list, a members-only list that presently includes more than 200 subscribers from about 20 countries. The mailing list is normally used for circulating open positions, calls for reviewers, calls for papers and invited session proposals, as well as for sharing advertisements on the activities of the CSS. The mailing list is also the vehicle for our newsletter, published quarterly, before and after the main conferences (CDC and ACC), and collecting announcements or reports on specific events organized at the conferences, reports on member activities, and other general information. A parallel means is the committee's Web site (http://www.diiga.univpm.it/wic/).

Nonetheless, the effective use of the mailing list is simultaneously a cause and effect of the need for meeting people, talking with them, generating new ideas, and organizing new events. A key role is played by the CSS-sponsored semiannual meetings

of the committee, organized at the CDCs and ACCs. Those meetings are open to all women attending the conference and are often focused on specific issues of interest to the community. Recently discussed topics include international research programs, funding opportunities, and qualifications and the process for becoming an IEEE Fellow. Moreover, since we believe that conference attendance is an essential aspect of scientific research, the committee provides a roommate location service at the main conferences, with the aim of reducing the costs of attending a conference, thus encouraging wider participation.

In a few words, by using very simple means and heavily relying on volunteer work and dedication, the Women in Control Committee aims at promoting visibility, networking, professional skills, and career development of women researchers in the field of control systems, to create for any woman the opportunities of identifying, cultivating, and promoting her talent.

Q. **What are some of your interests outside of research and teaching?**

Elena: Well, I should acknowledge that my work and the related activities do not concede much spare time to me. To quote a colleague, and dear friend of mine, in our job there are things that we *must* do and things that we *like* to do. The time left over is close to zero! Nonetheless, I do my best to cultivate a couple of extra interests.

As to the first, I am a member of a cultural association hosted by the Bologna Center of Johns Hopkins University. This association allows me to attend seminars given by international experts on topical issues in economics, politics, and sociology. Although I am, and I am afraid I will always be, an absolute beginner on those subjects, I have the feeling that it may be quite helpful to figure out how some processes work in the world where we live.

The second interest is indeed my favorite pastime and still has to do with project and design. In fact, I like reading books of architecture, residential design in particular. My original motivation was to collect information and suggestions to succeed in my "creative homemaking." However, while doing that, I saw how many inspiring issues are connected to the design of a home, from protection of the environment to health and happiness of those who live in it. Hence, I have become more and more interested in the subject.

Q. Thank you for speaking with *CSM!*

Elena: You're most welcome.

EYAD ABED

Q. What are some of your responsibilities as vice president for financial activities of CSS?

Eyad: As I learned from my immediate predecessor in this position, Jay Farrell, my role is to implement financial parameters approved by the Society's Executive Committee and Board of Governors. I work with IEEE headquarters to determine the Society's budget each year, to update the budget based on changes needed during the year, and to monitor the Society's overall financial condition. I also process reimbursement requests of Society volunteers and officers, and I prepare contracts with and process invoices from our vendors. Besides these routine activities, the vice president for financial activities is also a member of the Executive Committee and the Board of Governors and participates as a voting member in all of the deliberations of these committees.

Q. Please talk about some of the activities of the Institute for Systems Research at the University of Maryland.

Eyad: The Institute for Systems Research (ISR) is a cross-disciplinary research and education organization at the University of Maryland, College Park. At Maryland, the designation of a unit as an institute carries particular significance, as opposed to a center or laboratory for example. When the ISR was founded as one of the first group of National Science Foundation Engineering Research Centers (NSF ERC) in 1985, it was named the Systems Research Center and was housed jointly at Maryland and Harvard. However, when permanent institute status (and the base budget that comes with that status) was awarded by the State of Maryland in 1992, the name was changed to the Institute for Systems Research. So, after the ISR "graduated" from the NSF ERC program in 1996, the base budget allowed us to maintain the significant infrastructure that had been built. Although the ISR includes faculty from throughout the campus, for administrative purposes it falls under the office of the dean of Engineering.

The ISR brings faculty together from a variety of complementary disciplines to work on challenging research projects in systems. Of course, "systems" is a broad designation, and indeed the work of our faculty, students, and research staff covers a spectrum of research areas. Many universities promote cross-disciplinary research. At the ISR, it is our bread and butter. Let me give you a few examples illustrating the diversity of our programs. The ISR is home to HyNet, the Maryland Hybrid Networks Center, which is focused on hybrid communication network design and applications of communication networks to problems ranging from health care to military communications. HyNet is headed by John Baras, who also was the founding director of the ISR. We also house several laboratories focused on neuroscience and on building engineering systems that take advantage of lessons learned from biology and neuroscience. Faculty members participating in this collaborative effort come from departments throughout campus, including electrical engineering, biology, and psychology. In his Bode Lecture at the CDC last year, P.S. Krishnaprasad presented results from one of the neuroscience research projects, an interesting project that studies the behavior of the echolocating bat using a combination of experimental neuroscience combined with control theory. Finally, ISR also houses the head office of NEXTOR, the National Center of Excellence for Aviation Operations Research. This center, which is funded by the Federal Aviation Administration, is headed by Michael Ball, an ISR faculty member whose tenure resides in the business school. In addition to these and many other collaborative research projects, we also run a masters program in systems engineering as well as an NSF Research Experiences for Undergraduates program, and we have a vibrant industrial partners program.

Q. Your research has ranged from spacecraft control to power systems. What are some of the underlying ideas that you've pursued across these applications?

Eyad: There have indeed been a few underlying ideas and methodologies that are common to much of my research work. One thread of my research has been on developing theory and pursuing applications related to bifurcation control. This research has to do with control of nonlinear systems undergoing bifurcations from their usual operating condition to a new operating mode. Bifurcations happen when stability of the operating condition is lost due to parameter variations and is thus a very common phenomenon in systems of all types. The consequences of bifurcation can be as tame as an oscillating voltage level or as severe as an airplane entering a spin or a runaway reaction in a chemical plant. Another area of my research has dealt with singularly perturbed systems, which is a way to represent systems whose dynamics can be viewed approximately as consisting of

coupled slow and fast dynamics. I have also worked on several stability and dynamics problems for general linear systems, including robust stability and selective modal analysis.

Q. **What unsolved—open—research problem would you find most satisfying to conquer?**

Eyad: I believe that there are fundamental problems in the monitoring of uncertain systems that remain to be stated in a precise fashion, let alone being resolved. Control theorists have spent a great deal of effort on all types of control issues including robust control for uncertain systems. However, I don't think we have done anywhere near enough research on developing a theory (or theories) of system monitoring. Since real systems can be very difficult to model accurately and often involve parameters and dynamics that cannot be known, it is important to be able to make conclusions about the health of a system from online measurements despite uncertainty in the system model. Because of my past research on controlling systems undergoing bifurcation, I am interested in being able to determine from online measurements alone when a system is vulnerable to loss of stability and bifurcation. System identification tools focus on identifying a model, but that task can be both very difficult (especially for large systems) and is more than is required if we simply want to know whether a system is on the edge of losing its stability. This problem is an intriguing one and is important in many applications. For example, it would be wonderful to have a measurement-based stability monitoring tool for electric power networks. With such a tool, we would have some warning before a power system loses stability, rather than just having the less-powerful tools of today that can only detect an instability after it has begun. Instability detection at that stage is likely to be too little too late.

Q. **What are some of your current control projects?**

Eyad: I am working on power system monitoring, control of supercavi-

Eyad Abed with (from left) his daughters Maysoon and Mona and his wife Wendy (right). This picture was taken in Sabie, South Africa, near Kruger National Park in August 2007, just before the IFAC NOLCOS Symposium. Eyad is a professor in the Department of Electrical and Computer Engineering and Director of the Institute for Systems Research at the University of Maryland.

tating vehicles, modeling and stability analysis of rumor propagation dynamics, engineered swarms, control applications to accelerators, and applications of a new definition we have recently introduced for modal participation factors for linear systems.

Q. **What do you see as the key challenges of our field as a whole?**

Eyad: Our field has been very successful and has contributed greatly to technology and to society. However, most nonexperts have little knowledge that there even exists a coherent field of control systems. So I see as a key challenge achieving wide visibility for our field. Second, we should develop programs that introduce control to undergraduates early in their studies and give them an opportunity to consider specializing in control. With our interconnected world these days, we should reach out to the next generation of control scientists and engineers and assist and motivate them as much as possible. Third, I think that our field can contribute to modeling, analysis, and control of systems that involve human interactions. Finally, a key challenge for our field is to develop new techniques that facilitate uncertain system control and monitoring based on limited online measured data.

Q. **Was there a single instructor who motivated your interest in systems and control?**

Eyad: Actually, I was introduced to the field of systems and control not by an instructor but by a research scientist whom I had met somewhat serendipitously. This research scientist, Esteban (Steve) Hnyilicza, had done his graduate work with the late Fred Schweppe. He was working at the MIT Energy Laboratory on optimal stockpiling of petroleum reserves. After talking to me for a few minutes about my interests, he suggested that I spend some time in the library reading about control theory. So I did just that, and I found control to be an area that involved challenging mathematics as well as fascinating applications. I was immediately sold on the field and have been in the field ever since. I am still in touch with Steve, who is now a research professor of economics and a consultant in his native Lima, Peru.

Q. **Do you have any comments for those considering an academic career?**

Eyad: Sometimes I think about what's involved for a recent Ph.D. graduate to succeed in academia today, and it seems to be a daunting proposition. Success as a faculty

member requires a high degree of motivation and interest in one's chosen field, lots of hard work, and some degree of luck. I don't think these ingredients for success as a faculty member have changed since I joined Maryland as an assistant professor in 1983. However, some things have of course changed. Chief among these, I think, is the increased difficulty of achieving funding for one's research. Much of the research funding available today targets specific government needs, which are often short term. There are fewer opportunities for beginning faculty members to receive significant funding for topics of their own choosing that, while interesting, may not fit well into a stated government need. Despite the difficulties, however, I feel that I would still be interested in pursuing an academic career if I were starting out now. There are still significant opportunities in academia for innovation and success.

Q. **What trends do you see in control education, especially at the undergraduate level?**

Eyad: It's encouraging to see improvements in control laboratories, and it's also very helpful that project courses of various types have been introduced that have a major control component. The ever-increasing role of software tools in control education is in general positive, as long as students don't rely so heavily on the software that they fail to master the theory and techniques. One opportunity that software offers is to introduce students to a variety of issues that would otherwise be too advanced at the undergraduate level.

Q. **What are some of your hobbies and interests outside of professional activities?**

Eyad: I enjoy travel and learning about world cultures. I also enjoy reading, plays, classic movies, music, chess, and billiards.

Q. **Thank you for speaking with *CSM*!**

Eyad: Thank you!

2008 CSS Fellows

Election to the grade of IEEE Fellow acknowledges outstanding contributions and exceptional professional distinction. We are pleased to present the IEEE Control Systems Society members who have been accorded this honor for 2008.

Bassam Bamieh
University of California,
Santa Barbara

For contributions to robust, sampled-data, and distributed control.

Bassam Bamieh received the B.S. in electrical engineering and physics from Valparaiso University, Indiana, in 1983 and the M.Sc. and Ph.D. in electrical and computer engineering from Rice University, Houston, Texas, in 1986 and 1992, respectively. During 1991–1998, he was with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. He is currently a professor in the Mechanical Engineering Department at the University of California at Santa Barbara. His research interests are in distributed control and dynamical systems, shear flow turbulence modeling and control, quantum control, and thermoacoustic energy conversion devices. He is a Fellow of the IEEE, an IEEE Control Systems Society Distinguished Lecturer, a recipient of the IEEE Control Systems Society George S. Axelby Outstanding Paper Award, the AACC Hugo Schuck best paper award, and a National Science Foundation CAREER award.

Francis (Frank) J. Doyle III
University of California,
Santa Barbara

For contributions to nonlinear process control and analysis for biological systems.

Francis J. Doyle III is the associate director of the Army Institute for Collaborative Biotechnologies and the associate dean for research at the University of California, Santa Barbara. He holds the Duncan and Suzanne Mellichamp Chair in Process Control in the Department of Chemical Engineering as well as appointments in the Electrical Engineering Department and the Biomolecular Science and Engineering Program. He received his B.S.E. from Princeton (1985), C.P.G.S. from Cambridge (1986), and Ph.D. from Caltech (1991), all in chemical engineering. Previously, he held appointments at Purdue University and the University of Delaware, DuPont, Weyerhaeuser, and Stuttgart University. He is the recipient of several research and teaching awards, including the NSF National Young Investigator, ONR Young Investigator, Humboldt Research Fellowship, the Purdue Potter Award, and the ASEE Ray Fahien Award. He is currently the editor-in-chief of *IEEE Transactions on Control Systems Technology* and is an associate editor for the *Journal of Process Control, SIAM Journal on Applied Dynamical Systems,* and *Royal Society's Interface*. In 2005, he was awarded the Computing in Chemical Engineering Award from the American Institute of

member requires a high degree of motivation and interest in one's chosen field, lots of hard work, and some degree of luck. I don't think these ingredients for success as a faculty member have changed since I joined Maryland as an assistant professor in 1983. However, some things have of course changed. Chief among these, I think, is the increased difficulty of achieving funding for one's research. Much of the research funding available today targets specific government needs, which are often short term. There are fewer opportunities for beginning faculty members to receive significant funding for topics of their own choosing that, while interesting, may not fit well into a stated government need. Despite the difficulties, however, I feel that I would still be interested in pursuing an academic career if I were starting out now. There are still significant opportunities in academia for innovation and success.

Q. **What trends do you see in control education, especially at the undergraduate level?**

Eyad: It's encouraging to see improvements in control laboratories, and it's also very helpful that project courses of various types have been introduced that have a major control component. The ever-increasing role of software tools in control education is in general positive, as long as students don't rely so heavily on the software that they fail to master the theory and techniques. One opportunity that software offers is to introduce students to a variety of issues that would otherwise be too advanced at the undergraduate level.

Q. **What are some of your hobbies and interests outside of professional activities?**

Eyad: I enjoy travel and learning about world cultures. I also enjoy reading, plays, classic movies, music, chess, and billiards.

Q. **Thank you for speaking with *CSM*!**

Eyad: Thank you!

2008 CSS Fellows

Election to the grade of IEEE Fellow acknowledges outstanding contributions and exceptional professional distinction. We are pleased to present the IEEE Control Systems Society members who have been accorded this honor for 2008.

Bassam Bamieh
University of California, Santa Barbara

For contributions to robust, sampled-data, and distributed control.

Bassam Bamieh received the B.S. in electrical engineering and physics from Valparaiso University, Indiana, in 1983 and the M.Sc. and Ph.D. in electrical and computer engineering from Rice University, Houston, Texas, in 1986 and 1992, respectively. During 1991–1998, he was with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. He is currently a professor in the Mechanical Engineering Department at the University of California at Santa Barbara. His research interests are in distributed control and dynamical systems, shear flow turbulence modeling and control, quantum control, and thermoacoustic energy conversion devices. He is a Fellow of the IEEE, an IEEE Control Systems Society Distinguished Lecturer, a recipient of the IEEE Control Systems Society George S. Axelby Outstanding Paper Award, the AACC Hugo Schuck best paper award, and a National Science Foundation CAREER award.

Francis (Frank) J. Doyle III
University of California, Santa Barbara

For contributions to nonlinear process control and analysis for biological systems.

Francis J. Doyle III is the associate director of the Army Institute for Collaborative Biotechnologies and the associate dean for research at the University of California, Santa Barbara. He holds the Duncan and Suzanne Mellichamp Chair in Process Control in the Department of Chemical Engineering as well as appointments in the Electrical Engineering Department and the Biomolecular Science and Engineering Program. He received his B.S.E. from Princeton (1985), C.P.G.S. from Cambridge (1986), and Ph.D. from Caltech (1991), all in chemical engineering. Previously, he held appointments at Purdue University and the University of Delaware, DuPont, Weyerhaeuser, and Stuttgart University. He is the recipient of several research and teaching awards, including the NSF National Young Investigator, ONR Young Investigator, Humboldt Research Fellowship, the Purdue Potter Award, and the ASEE Ray Fahien Award. He is currently the editor-in-chief of *IEEE Transactions on Control Systems Technology* and is an associate editor for the *Journal of Process Control, SIAM Journal on Applied Dynamical Systems,* and *Royal Society's Interface*. In 2005, he was awarded the Computing in Chemical Engineering Award from the American Institute of

Chemical Engineers for work in systems biology. His research interests are in systems biology, network science, modeling and analysis of circadian rhythms, drug delivery for diabetes, model-based control, and control of particulate processes.

Geir E. Dullerud
University of Illinois at Urbana-Champaign

For contributions to sampled-data systems and robust control.

Geir E. Dullerud is currently professor of mechanical engineering at the University of Illinois, Urbana-Champaign (UIUC), and is also a research professor in the Coordinated Science Laboratory. He received the B.A.Sc. in engineering science in 1988 and the M.A.Sc. in electrical engineering in 1990, both from the University of Toronto, Canada. In 1994 he received the Ph.D. in engineering from the University of Cambridge, U.K. From 1996 to 1998 he was an assistant professor in applied mathematics at the University of Waterloo, Canada. During 1994 and 1995 he was a research fellow at the California Institute of Technology in the Control and Dynamical Systems Department. He has coauthored two books, *Control of Uncertain Sampled-Data Systems* and *A Course in Robust Control Theory*. His research interests include networks, complex and hybrid dynamical systems, and control of distributed robotic systems. He received the Xerox Award at UIUC in 2005 and the National Science Foundation CAREER Award in 1999.

Michael C. Fu
University of Maryland

For contributions to stochastic gradient estimation and simulation optimization.

Michael C. Fu is the Ralph J. Tyser Professor of Management Science in the Robert H. Smith School of Business at the University of Maryland, with joint appointments in the Institute for Systems Research and the Department of Electrical and Computer Engineering. He received degrees in mathematics and electrical engineering/computer science from MIT and a Ph.D. in applied mathematics from Harvard University. His research interests include simulation, optimization, and applied probability, with applications in supply chain management and financial engineering. He has published four books, *Monte Carlo: Gradient Estimation and Optimization Applications, Simulation-Based Algorithms for Markov Decision Processes, Perspectives in Operations Research*, and *Advances in Mathematical Finance*. He has published over 100 journal articles, book chapters, and conference proceedings. His awards include the Allen J. Krowe Award for Teaching Excellence (1995), the University of Maryland Distinguished Scholar-Teacher (2004), the Institute for Systems Research Outstanding Systems Engineering Faculty Award (2002), the IIE Operations Research Division Award (1999), and the IIE Transactions Best Paper Award (1998). He currently serves as stochastic models and simulation department editor for *Management Science* and associate editor for *Mathematics of Operations Research*. He was a simulation area editor of *Operations Research* (2000–2005) and served on the program committee for the 2007 IEEE Conference on Decision and Control. He is a fellow of the Institute for Operations Research and the Management Sciences.

Fred Hadaegh
Jet Propulsion laboratory

For contributions to autonomous guidance and control systems for space vehicles and multispacecraft formations.

Fred Hadaegh received the B.S. and M.S. in electrical engineering from the University of Texas at Austin and the Ph.D. in electrical engineering from the University of Southern California, Los Angeles. He joined the Jet Propulsion Laboratory (JPL) in 1984, where he is a senior research scientist, the supervisor of the guidance and control analysis group, and manager of the Distributed Spacecraft Technology Program. He is a fellow of the American Institute of Aeronautics and Astronautics (AIAA). He is a recipient of NASA's Exceptional Service Medal, NASA's Exceptional Achievement Medal, and the Jet Propulsion Laboratory's Award of Excellence for his contributions to *Flight Validation of Autonomous Rendezvous in Earth Orbit*. His research contributions are in mathematical modeling of uncertain systems, system identification, control of large space structures, and optimal estimation and autonomous control of formation flying of space systems.

João P. Hespanha,
University of California, Santa Barbara

For contributions to stability techniques for switched and hybrid systems.

João P. Hespanha received the licenciatura in electrical and computer engineering from the Instituto Superior Técnico, Lisbon, Portugal in 1991 and the Ph.D. in electrical engineering and applied science from Yale University, New Haven, Connecticut, in 1998. From 1999 to 2001, he was an assistant professor at the University of Southern California, Los Angeles. In 2002 he joined the University of California, Santa Barbara, where he is a professor in the Department of Electrical and Computer Engineering. He is associate director for the Center for Control, Dynamical Systems,

and Computation, vice-chair of the Department of Electrical and Computer Engineering, and a member of the executive committee for the Institute for Collaborative Biotechnologies. From 2004–2007 he was an associate editor for *IEEE Transactions on Automatic Control.*

He is the recipient of Yale University's Henry Prentiss Becton Graduate Prize for exceptional achievement in research in engineering and applied science, a National Science Foundation CAREER Award, the 2005 best paper award at the Second International Conference on Intelligent Sensing and Information Processing, the 2005 *Automatica* Theory/Methodology best paper prize, and the 2006 George S. Axelby Outstanding Paper Award. He is an IEEE Distinguished Lecturer. His research interests include hybrid and switched systems, modeling and control of communication networks, distributed control of communication, the use of vision in feedback control, and stochastic modeling of regulatory systems in biology.

Zhong-Ping Jiang
Polytechnic University
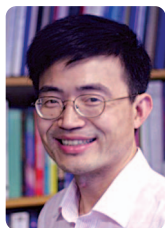
For contributions to nonlinear control theory and underactuated mechanical systems.

Zhong-Ping Jiang received the B.Sc. in mathematics from the University of Wuhan, China, in 1988, the M.Sc. in statistics from the Université de Paris-sud, France, in 1989, and the Ph.D. in automatic control and mathematics from the École des Mines de Paris, France, in 1993. He is a professor in the Department of Electrical and Computer Engineering at Polytechnic University, New York. He has published in the areas of stability theory, robust and adaptive nonlinear control, underactuated mechanical systems, congestion control, and wireless networks. His research interests include cooperative

control for multiagent systems and human dynamic decision making. He has served in editorial positions for the *International Journal of Robust and Nonlinear Control, Systems and Control Letters, IEEE Transactions on Automatic Control*, and the *European Journal of Control*. He is a recipient of the Queen Elizabeth II Fellowship Award from the Australian Research Council, the CAREER Award from the U.S. National Science Foundation, and the Young Investigator Award from the NSF of China.

Wei Kang
Naval Postgraduate School

For contributions to industrial applications of nonlinear control systems.

Wei Kang received the B.S. and M.S. from Nankai University, China, in 1982 and 1985, respectively, and the Ph.D. from the University of California at Davis in 1991, all in mathematics. He previously held a faculty position at Washington University in St. Louis. Since 1994 he has been a professor of applied mathematics at the U.S. Naval Postgraduate School. His research interests include computational mathematics in control design, bifurcations and normal forms, H_∞ control, manufacturing and process control, autonomous vehicles, and space systems. He has served as an associate editor of *IEEE Transactions on Automatic Control, Automatica,* and the *Journal of Control Theory and Applications*. He was a plenary speaker at the Seventh SIAM Conference on Control and Applications and the First IFAC Conference on Control and Analysis of Chaotic Systems. He is the recipient of the Best Paper Award of the Sixth International Conference on Control, Automation, Robotics and Vision held in Singapore in 2000, and the Carl E. and Jessie W. Menneken Faculty Award from NPS Foundation.

Ilya V. Kolmanovsky
Ford Motor Company

For contributions to nonlinear control of automotive powertrains.

Ilya V. Kolmanovsky received an undergraduate degree from the Moscow Aviation Institute in Russia. He received the M.S. and Ph.D. in aerospace engineering and the M.A. in mathematics from the University of Michigan, Ann Arbor, in 1993, 1995, and 1995, respectively. He is a technical leader at Ford Motor Company in Dearborn, Michigan, where he has overseen the development of powertrain controls for advanced gasoline and diesel engines. He is an adjunct research scientist at the University of Michigan. He has published over 180 journal and conference articles and is an inventor on over 68 United States patents. He has formed and participated in industry-academia partnerships to facilitate synergistic advances in control theory and applications. He has been an associate editor of the IEEE Control Systems Society Conference Editorial Board, *IEEE Transactions on Control Systems Technology,* and *IEEE Transactions on Automatic Control*. He has been a chair of the IEEE Control Systems Society Technical Committee on Automotive Control, and he served as a vice chair of the 2006 IEEE CDC. He was a recipient of the Donald P. Eckman Award of American Automatic Control Council and the IEEE Transactions on Control Systems Technology Outstanding Paper Award.

Dragan Nesic
University of Melbourne

For contributions to the analysis and control of networked nonlinear sampled-data systems.

Dragan Nesic is a professor in the Department of Electrical and

Electronic Engineering (DEEE) at the University of Melbourne, Australia. He received the B.E. in mechanical engineering from the University of Belgrade, Yugoslavia, in 1990, and the Ph.D. in systems engineering from the Australian National University, Canberra, in 1997. From 1997 to 1999 he held several postdoctoral positions. His research interests include networked control systems, discrete-time, sampled-data, and continuous-time nonlinear control systems, input-to-state stability, extremum seeking control, applications of symbolic computation in control theory, and hybrid control systems. He was awarded a Humboldt Research Fellowship (2003) by the Alexander von Humboldt Foundation and an Australian Professorial Fellowship (2004–2009) by the Australian Research Council. He is a fellow of IEAust. He is currently a Distinguished Lecturer of IEEE Control Systems Society as well as an associate editor for *Automatica, IEEE Transactions on Automatic, Systems and Control Letters*, and the *European Journal of Control*.

Andrew Packard
University of California, Berkeley

For contributions to robust control.

Andrew Packard joined the University of California, Berkeley, Mechanical Engineering Department in 1990. He teaches courses in control, applied engineering, mathematics, introductory scientific programming, dynamics, and acoustics. In 1988, he was a postdoctoral researcher and lecturer at Caltech, and in 1989 he was an assistant professor at the University of California, Santa Barbara, in electrical and computer engineering. His research interests include computational methods for quantitative estimates of nonlinear system behavior, and combining sum-of-squares optimization with simulation to discover Lyapunov/storage functions that certify bounds on possible behavior. His current focus is on algorithms and pragmatic approaches to large-scale, distributed system identification problems. He is a recipient of the 1995 University of California, Berkeley, Distinguished Teaching Award, the 1995 Eckman Award, and the 2005 IEEE Control System Technology Award. He is a coauthor of both the *muTools and Robust Control toolboxes* (Mathworks). The Meyersound X-10 loudspeaker utilizes feedback control circuitry developed by his research group.

Paul M.J. Van den Hof
Delft University of Technology

For contributions to system identification for control systems.

Paul M.J. Van den Hof received the M.Sc. and Ph.D. from the Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands, in 1982 and 1989, respectively. In 1986 he joined the Mechanical Engineering Systems and Control Group of Delft University of Technology, The Netherlands. In 1999 he was appointed professor in the Department of Applied Physics. He is currently the director of the Delft Center for Systems and Control. Since 2005 he has served as the scientific director of the Dutch Institute of Systems and Control. His research concentrates on system identification and identification for control, including aspects of closed-loop experiments, uncertainty modeling and parameterization issues, in particular by employing orthonormal basis functions. He was the general chair for the 13th IFAC Symposium on System Identification, Rotterdam, The Netherlands. He has been a member of the editorial board of *Automatica*, the IFAC Council, the Administrative Council of the European Union Control Association EUCA, and the Board of Governors of IEEE Control System Society. In 2007 he was elected as an IFAC fellow.

Kevin A. Wise
The Boeing Company

For contributions to adaptive and optimal control of unmanned aerospace systems.

Kevin A. Wise is a senior technical fellow for advanced flight controls at the Boeing Company. He received the B.S., M.S., and Ph.D. in mechanical engineering from the University of Illinois in 1980, 1982, and 1987, respectively. He has developed vehicle management systems, flight control systems, and control system design tools and processes for advanced unmanned aircraft and weapon system programs. Highlights include the High Altitude Long Endurance (HALE) UAV (2004–2005), Laser-JDAM (in production), X-45A J-UCAS (1999–2004), X-36 RESTORE (1996–1999), miniaturized munitions technology demonstrator (1996–1999), fourth generation ejection seat (first supersonic ejection 1996), JDAM (in production), Army short range UAV (1990–1992), and the HAVE SLICK missile (1986–1991), and model reference adaptive control for the MK-82 Laser-JDAM and JDAM weapon systems. His research interests include autonomous systems and decision making, aircraft and missile dynamics and control, robust adaptive control of linear and nonlinear systems, and robustness theory for parametric and dynamic uncertainties. He has authored more than 60 technical articles and taught 17 workshops. He teaches control theory at Washington University, Southern Illinois University at Edwardsville, and at the University of Missouri Science and Technology.

MEET THE NEW *CSM* ASSOCIATE EDITORS

Penina Axelrad is a professor and acting chair of the Aerospace Engineering Sciences Department at the University of Colorado (CU), Boulder. She received the B.S. and M.S. in aeronautical and astronautical engineering, with an emphasis on avionics from MIT in 1985 and 1986, respectively, and the Ph.D. in aeronautics and astronautics from Stanford University in 1991. In 1991–1992 she was a member of the technical staff at Stanford Telecommunications. In 1992 she joined the faculty at CU. Her research interests include technology and algorithms for GPS-based orbit and attitude determination for spacecraft, navigation of aircraft and land vehicles, multipath characterization and correction for spacecraft, aircraft, and ground reference stations, and remote sensing and terrain awareness using GPS-based bistatic radar. She is a coauthor of 35 journal publications and 80 conference papers. She is a fellow of the Institute of Navigation and of the American Institute of Aeronautics and Astronautics (AIAA), a Senior Member of IEEE, and a member of Sigma Xi. She received the 1996 Lawrence Sperry Award from the AIAA and the 2003 Tycho Brahe Award from the ION.

Silvia Ferrari is an assistant professor of mechanical engineering and materials science at Duke University, where she directs the Laboratory for Intelligent Systems and Controls Her research interests include robust adaptive control of aircraft, learning and approximate dynamic programming, and optimal control of mobile sensor networks. She received the B.S. from Embry-Riddle Aeronautical University and the M.A. and Ph.D. from Princeton University. She is a Member of the IEEE, ASME, SPIE, and AIAA. She is the recipient of the ONR Young Investigator Award (2004), the NSF CAREER Award (2005), and the Presidential Early Career Award for Scientists and Engineers (2006).

Rafael Fierro received the M.Sc. in control engineering from the University of Bradford, U.K., in 1990 and the Ph.D. in electrical engineering from the University of Texas at Arlington in 1997. He joined the University of New Mexico in 2007, where he is currently an associate professor in the Department of Electrical and Computer Engineering. He previously held a postdoctoral appointment with the GRASP Lab at the University of Pennsylvania and a faculty position with the Department of Electrical and Computer Engineering at Oklahoma State University. His research interests include hybrid and embedded systems, optimization-based cooperative control, and heterogeneous multivehicle coordination. His current research focuses on developing systematic methodologies for the analysis, design, and implementation of cooperative dynamic networks. At the University of New Mexico he directs the Multi-Agent, Robotics, Hybrid and Embedded Systems Laboratory. He and his students are developing robotic games to increase the interest of K–12 students and their appreciation for science, math, and engineering. He was the recipient of a Fulbright scholarship and a 2004 National Science Foundation CAREER award. He is an associate editor for *IEEE Transactions on Automation Science and Engineering* and the *Journal of Intelligent and Robotics Systems*.

Kathryn E. Johnson received the B.S. in electrical engineering from Clark-son University in 2000 and the M.S. and Ph.D. in electrical engineering from the University of Colorado in 2002 and 2004, respectively. In 2005, she completed a postdoctoral position researching adaptive control of variable-speed wind turbines at the National Renewable Energy Laboratory's National Wind Technology Center. She is currently the Clare Boothe Luce assistant professor at the Colorado School of Mines in the Division of Engineering. Her research interests focus on control systems and control applications. Her wind energy control projects include assessing the value of continuously variable transmissions in a small wind turbine, determining the design-driving load cases in utility-scale wind turbines, and improving the energy capture of wind farms using coordinated turbine control. She enjoys cross-country and telemark skiing, mountain biking, and volunteering with the Rocky Mountain Rescue Group.

Eric Klavins received the Ph.D. in computer science and engineering from the University of Michigan in 2001. From 2001 to 2003, he was a postdoctoral scholar at the California Institute of Technology. Since 2003, he has been an assistant professor in the Department of Electrical Engineering at the University of Washington, Seattle. In 2004 he received an NSF CAREER award to study robotic self-assembly. His research interests include control engineering, dynamical systems, and distributed algorithms applied to distributed robotics, molecular devices, and systems biology.

John M. Watkins received the B.S. in electrical engineering from the University of Nebraska-Lincoln in 1989 and

the M.S. and Ph.D. in electrical engineering from the Ohio State University in 1991 and 1995, respectively. From 1995 to 2004 he was a faculty member at the U.S. Naval Academy. Since 2004 he has been an associate professor in the Department of Electrical and Computer Engineering at Wichita State University. He served as vice-chair for interactive papers of the 2007 CDC, registration chair of the 2006 ACC, and finance chair of the 2002 ACC. Presently, he is Student Affairs chair for the 2008 IEEE Multiconference on Systems and Control. He has been an associate editor of the Conference Editorial Board of the IEEE Control Systems Society since 1999. He was a NASA-ASEE Summer Faculty Fellow in 1999 and 2000. In 2007 he received the Wallace Excellence in Teaching Award. His research interests include time-delay systems, fault detection and isolation, active magnetic bearings, active noise control, mobile sensor networks, and spacecraft dynamics and control. (CSS)

---

rates. One TCP variant, TCP-Reno, increments sending rates by a fixed amount for each positive ack, and approximately halves the rate on receipt of a negative ack. Said another way, we have the control law

$$\frac{d}{dt} \text{ sending rate}$$
$$= \begin{cases} \text{constant,} & \text{if positive ack} \\ -\frac{1}{2} \text{ sending rate,} & \text{if negative ack.} \end{cases}$$

Consequently, the amount of data injected into the Internet increases linearly until congestion is sensed (by means of a negative ack) and then backs off exponentially.

Up to this point, we have presented a simplistic view of feedback control mechanisms in the Internet. In reality, the situation is much more complex, providing challenges to systems modeling, dynamical analysis, and control design; see Figure 2.

As a dynamic system, the Internet is an example of a hybrid system wherein actions are triggered by events; for example, the decrease in sending rates is triggered by congestion events. The Internet is also an example of a large-scale decentralized control system, where many data senders inject data into Internet datapipes, each bound for a specific datareceiver; where each sending-rate protocol reacts only to congestion signals generated by its data.

The existence of time delays in the Internet also plays a crucial role in feedback control, a situation that every student of feedback control can appreciate. These time delays are comprised of the propagation delays associated with data traveling through data pipes at the finite speed of light as well as by queuing delays encountered by data moving through the Internet's congested entrance and exit ramps. As in any feedback control system, these delays restrict the gain of a feedback controller or, in the case of the Internet, restrict the aggressiveness with which the sending-rate protocols react to congestion signals. Finally, the whole notion of "controller implementation" or "sensor implementation" in the Internet is challenging, since it must be accomplished using the data itself and existing protocol structures.

In this discussion, I've given a high-level view of the role that feedback plays in keeping the Internet running smoothly. Many important details have been abstracted out. To probe further, we recommend the three sources below including the references contained therein.

REFERENCES

[1] J.F. Kurose and K.W. Ross, *Computer Networking: A Top-Down Approach*, 4th ed. Reading, MA: Addison-Wesley, 2008.
[2] R. Srikant, *The Mathematics of Internet Congestion Control*. New York: Birkhäuser, 2003.
[3] S.H. Low, F. Paganini, and J.C. Doyle, "Internet congestion control," *IEEE Contr. Syst. Mag.*, vol. 22, no. 1, pp. 28–43, Feb. 2002.

AUTHOR INFORMATION

Kris Hollot received the Ph.D. in electrical engineering from the University of Rochester, New York, in 1984. Since 1984, he has been with the Department of Electrical and Computer Engineering at the University of Massachusetts, Amherst. He has served as associate editor for several control journals, and his research interests are in the theory and application of feedback control. He received the National Science Foundation Presidential Young Investigator Award in 1988. (CSS)

Modeling and Control of Adhesion Force in Railway Rolling Stocks

SUNG HWAN PARK, JONG SHIK KIM, JEONG JU CHOI, and HIRO-O YAMAZAKI

## ADAPTIVE SLIDING MODE CONTROL FOR THE DESIRED WHEEL SLIP

Studies of braking mechanisms of railway rolling stocks focus on the adhesion force, which is the tractive friction force that occurs between the rail and the wheel [1]–[3]. During braking, the wheel always slips on the rail. The adhesion force increases or decreases according to the slip ratio, which is the difference between the velocity of the rolling stocks and the tangential velocity of each wheel of the rolling stocks normalized with respect to the velocity of the rolling stocks. A nonzero slip ratio always occurs when the brake caliper holds the brake disk and thus the tangential velocity of the wheel so that the velocity of the wheel is lower than the velocity of the rolling stocks. Unless an automobile is skidding, the slip ratio for an automobile is always zero. In addition, the adhesion force decreases as the rail conditions change from dry to wet [3], [4]. Furthermore, since it is impossible to directly measure the adhesion force, the characteristics of the adhesion force must be inferred based on experiments [2], [3].

To maximize the adhesion force, it is essential to operate at the slip ratio at which the adhesion force is maximized. In addition, the slip ratio must not exceed a specified value determined to prevent too much wheel slip. Therefore, it is necessary to characterize the adhesion force through precise modeling.

To estimate the adhesion force, observer techniques are applied [5]. In addition, based on the estimated value,

wheel-slip brake control systems are designed [6], [7]. However, these control systems do not consider uncertainty such as randomness in the adhesion force between the rail and the wheel. To address this problem, a reference slip ratio generation algorithm is developed by using a disturbance observer to determine the desired slip ratio for maximum adhesion force [8]–[10]. Since uncertainty in the traveling resistance and the mass of the rolling stocks is not considered, the reference slip ratio, at which adhesion force is maximized, cannot always guarantee the desired wheel slip for good braking performance.

In this article, two models are developed for the adhesion force in railway rolling stocks. The first model is a static model based on a beam model, which is typically used
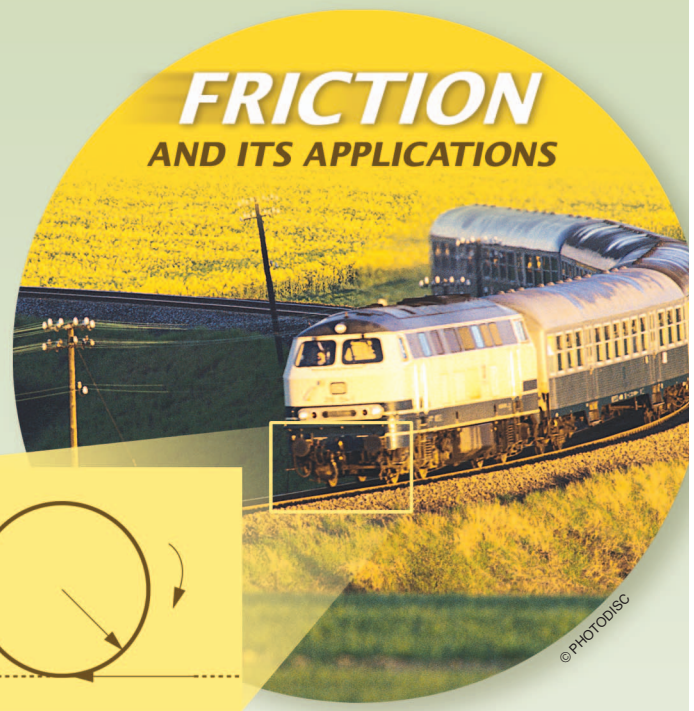
to model automobile tires. The second model is a dynamic model based on a bristle model, in which the friction interface between the rail and the wheel is modeled as contact between bristles [11]. The validity of the beam model and bristle model is verified through an adhesion test using a brake performance test rig.

We also develop wheel-slip brake control systems based on each friction model. One control system is a conventional proportional-integral (PI) control scheme, while the other is an adaptive sliding mode control (ASMC) scheme. The controller design process considers system uncertainties such as the traveling resistance, disturbance torque, and variation of the adhesion force according to the slip ratio and rail conditions. The mass of the rolling stocks is also considered as an uncertain parameter, and the adaptive law is based on Lyapunov stability theory. The performance and robustness of the PI and adaptive sliding mode wheel-slip brake control systems are evaluated through computer simulation.

WHEEL-SLIP MECHANISM FOR ROLLING STOCKS

To reduce braking distance, automobiles are fitted with an antilock braking system (ABS) [12], [13]. However, there is a relatively low adhesion force between the rail and the wheel in railway rolling stocks compared with automobiles. A wheel-slip control system, which is similar to the ABS for automobiles, is currently used in the brake system for railway rolling stocks.

The braking mechanism of the rolling stocks can be modeled by

$$F_a = \mu(\lambda)N, \tag{1}$$
$$\lambda = \frac{v - r\omega}{v}, \tag{2}$$

where F_a is the adhesion force, $\mu(\lambda)$ is the dimensionless adhesion coefficient, λ is the slip ratio, N is the normal force, v is the velocity of the rolling stocks, and ω are r the angular velocity and radius of each wheel of the rolling stocks, respectively. The velocity of the rolling stocks can be measured [14] or estimated [15]. The adhesion force F_a is the friction force that is orthogonal to the normal force. This force disturbs the motion of the rolling stocks desirably or undesirably according to the relative velocity between the rail and the wheel. The adhesion force F_a changes according to the variation of the adhesion coefficient $\mu(\lambda)$, which depends on the slip ratio λ, railway condition, axle load, and initial braking velocity, that is, the velocity at which the brake is applied. Figure 1 shows a typical shape of the adhesion coefficient $\mu(\lambda)$ according to the slip ratio λ and rail conditions.

To design a wheel-slip control system, it is useful to simplify the dynamics of the rolling stocks as a quarter model based on the assumption that the rolling stocks travel in the longitudinal direction without lateral motion, as shown in Figure 2. The equations of motion for the quarter model of the rolling stocks can be expressed as

$$J\dot{\omega} = -B\omega + T_a - T_b - T_d, \tag{3}$$
$$M\dot{v} = -F_a - F_r, \tag{4}$$

where B is the viscous friction torque coefficient between the brake pad and the wheel; $T_a = rF_a$ and T_b are the adhesion and brake torques, respectively; T_d is the disturbance torque due to the vibration of the brake caliper; J and r are the inertia and radius, respectively, of each wheel of the rolling stocks; and M and F_r are the mass and traveling resistance force of the rolling stocks, respectively.
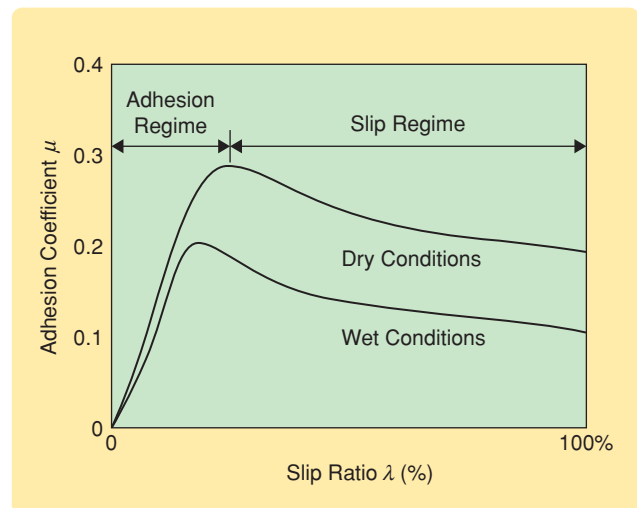
FIGURE 1 Typical shape of the adhesion coefficient according to the slip ratio and rail conditions. The adhesion coefficient increases in the adhesion regime, while the adhesion coefficient decreases in the slip regime according to the slip ratio. The properties of the adhesion force change greatly according to the variation of the adhesion coefficient, which depends on the slip ratio, rail conditions, axle load, and initial braking velocity.
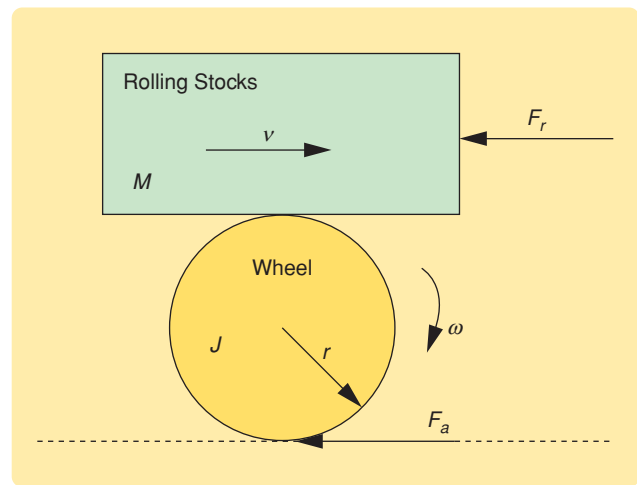
FIGURE 2 Quarter model of the rolling stocks. The dynamics of the rolling stocks are simplified by using a quarter model, which is based on the assumption that the rolling stocks travel in the longitudinal direction without lateral motion.

From (3) and (4), it can be seen that, to achieve sufficient adhesion force, a large brake torque T_b must be applied. When T_b is increased, however, the slip ratio increases, which causes the wheel to slip. When the wheel slips, it may develop a flat spot on the rolling surface. This flat spot affects the stability of the rolling stocks, the comfort of the passengers, and the life cycle of the rail and the wheel. To prevent this undesirable braking situation, a desired wheel-slip control is essential for the brake system of the rolling stocks.

In addition, the adhesion force between the wheel and the contact surface is dominated by the initial braking veloc-

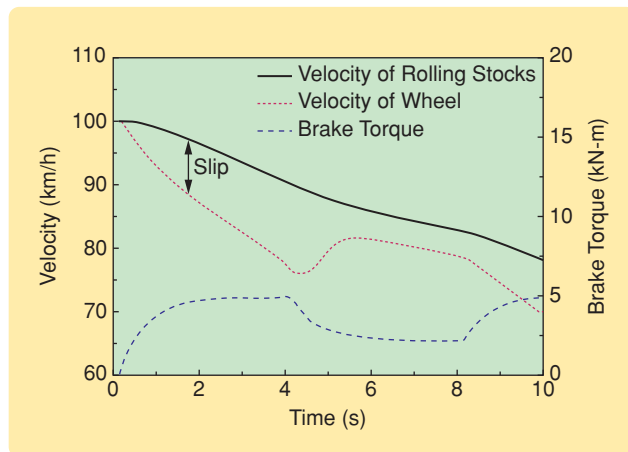**FIGURE 3** An example of a wheel-slip control mechanism based on the relationship between the slip ratio and braking performance. Railway systems usually monitor the rotation of each axle and compare the rotational speeds between pairs of axles. If a difference appears during braking, the brake is released or reapplied to those axles so that the reference slip ratio can be obtained.
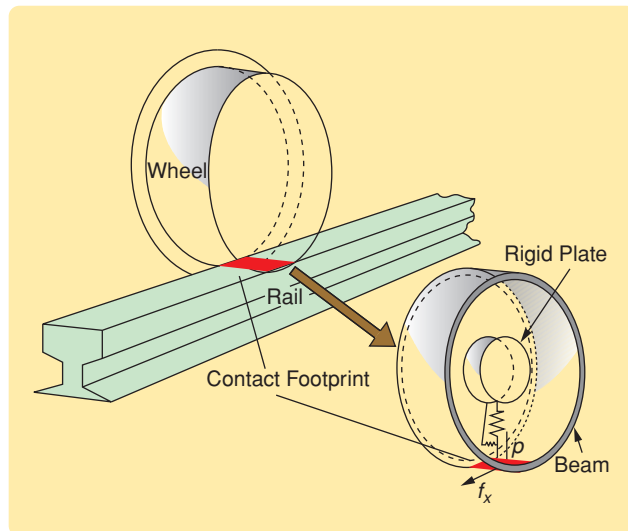

FIGURE 4 Simplified contact model for the rail and wheel. The beam model treats the wheel as a circular beam supported by springs. The variable p is the contact pressure between the rail and the wheel, while the variable f_x is the derivative of the adhesion force with respect to the displacement from the tip of the contact footprint in the longitudinal direction.

ity, as well as by the mass M and railway conditions. In the case of automobiles, which have rubber pneumatic tires, the maximum adhesion coefficient changes from 0.4 to 1 according to the road conditions and the materials of the contact surface [16], [17]. In the case of railway rolling stocks, where the contact between the wheel and the rail is that of steel on steel, the maximum adhesion force coefficient changes from approximately 0.1 to 0.4 according to the railway conditions and the materials of the contact surface [18]. Therefore, railway rolling stocks and automobiles have significantly different adhesion force coefficients because of different materials for the rolling and contact surfaces. However, the brake characteristics of railway rolling stocks [19], [20] and automobiles [13], [17] are similar.

According to adhesion theory [18], the maximum adhesion force occurs when the slip ratio is approximately between 0.1 and 0.4 in railway rolling stocks. Therefore, the slip ratio at which the maximum adhesion force is obtained is usually used as the reference slip ratio for the brake control system of the rolling stocks. Figure 3 shows an example of a wheel-slip control mechanism based on the relationship between the slip ratio and braking performance.

STATIC ADHESION FORCE MODEL BASED ON THE BEAM MODEL

To model the adhesion force as a function of the slip ratio, we consider the beam model, which reflects only the longitudinal adhesion force. Figure 4 shows a simplified contact model for the rail and wheel, where the beam model treats the wheel as a circular beam supported by springs. The contact footprint of an automobile tire is generally approximated as a rectangle by the beam model [21]. In a similar manner, the contact footprint between the rail and the wheel is approximated by a rectangle as shown in Figure 5.

The contact pressure p between the rail and the wheel at the displacement x_c from the tip of the contact footprint in the longitudinal direction is given by [21]

$$p = \frac{6N}{l^3 w}\left[\left(\frac{l}{2}\right)^2 - \left(x_c - \frac{l}{2}\right)^2\right], \qquad (5)$$

where N is the normal force and l and w are the length and width of the contact footprint, respectively. Figure 6 shows a typical distribution of the tangential force coefficient in a contact footprint [22]–[24].

In Figure 6, the variable f_x, which is the derivative of the adhesion force F_a with respect to the displacement x_c from the tip of the contact footprint, is given by

$$f_x = \begin{cases} C_x \lambda w x_c & \text{for } 0 \leq x_c \leq l_h, \\ \mu_d p & \text{for } l_h < x_c \leq l, \end{cases} \qquad (6)$$

where C_x is the modulus of transverse elasticity, l_h is the displacement from the tip of the contact footprint at which the

adhesion-force derivative f_x changes rapidly, and μ_d is the dynamic friction coefficient. In particular, μ_d is defined by

$$\mu_d = \mu_{\max} - \frac{a\lambda vl}{(l - l_h)}, \tag{7}$$

where $\mu_{\max}$ is the maximum adhesion coefficient, a is a constant that determines the dynamic friction coefficient in the slipping regime, and l_h is expressed as [21]

$$l_h = l\left(1 - \frac{K_x\lambda}{3\mu_{\max}N}\right), \tag{8}$$

where K_x is the traveling stiffness calculated by

$$K_x = \frac{1}{2}C_x l^2. \tag{9}$$

The wheel load, which is the normal force, is equal to the integrated value of the contact pressure between the rail and the wheel over the contact footprint. Therefore, the adhesion force F_a between the rail and the wheel can be calculated by integrating (6) over the length of the contact footprint and substituting (7) and (8) into (6), which is expressed as

$$F_a = \frac{1}{2}C_x\lambda wl^2\left(1 - \frac{K_x\lambda}{3\mu_{\max}N}\right)^2 + \frac{1}{2}K_x\lambda - \frac{3}{2}Na(v - r\omega)$$
$$- \frac{1}{2}\mu_{\max}N\left[1 - 3\frac{Na(v - r\omega)}{K_x\lambda}\right]\left[1 - \left(1 - \frac{2K_x\lambda}{3\mu_{\max}N}\right)^3\right]. \tag{10}$$

DYNAMIC ADHESION FORCE MODEL BASED ON BRISTLE CONTACT

As a dynamic adhesion force model, we consider the Dahl model given by [25]

$$\frac{dz}{dt} = \sigma - \frac{\alpha|\sigma|}{F_c}z, \tag{11}$$
$$F = \alpha z, \tag{12}$$

where z is the internal friction state, σ is the relative velocity, α is the stiffness coefficient, and F and F_c are the friction force and Coulomb friction force, respectively. Since the steady-state version of the Dahl model is equivalent to Coulomb friction, the Dahl model is a generalized model for Coulomb friction. However, the Dahl model does not capture either the Stribeck effect or stick-slip effects. In fact, the friction behavior of the adhesion force according to the relative velocity σ for railway rolling stocks exhibits the Stribeck effect, as shown in Figure 7. Therefore, the Dahl model is not suitable as an adhesion force model for railway rolling stocks.

However, the LuGre model [11], which is a generalized form of the Dahl model, can describe both the Stribeck effect and stick-slip effects. The LuGre model equations are given by

$$\frac{dz}{dt} = \sigma - \frac{\alpha|\sigma|}{g(\sigma)}z, \tag{13}$$
$$g(\sigma) = F_c + (F_s - F_c)e^{-(\sigma/v_s)^2}, \tag{14}$$
$$F = \alpha z + \alpha_1\dot{z} + \alpha_2\sigma, \tag{15}$$
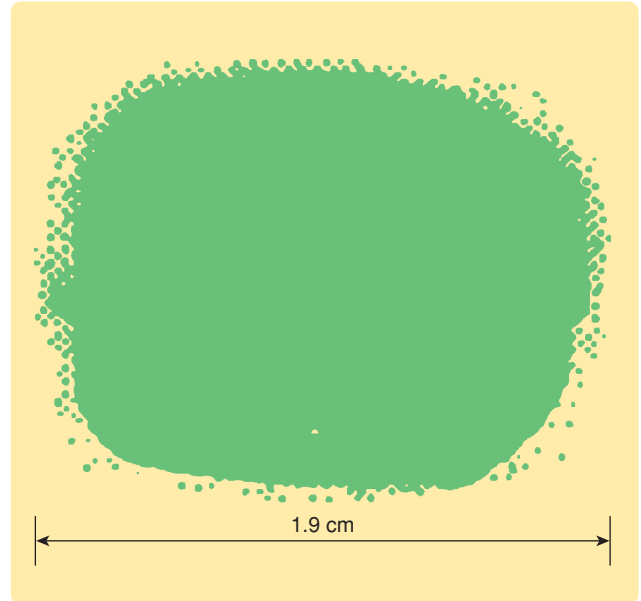
FIGURE 5 Contact footprint between the rail and the wheel. The shape of this footprint can be approximated by a rectangle. The length and width of the contact footprint are both approximately 1.9 cm. This value is used to calculate the contact pressure between the rail and the wheel as well as the adhesion force.
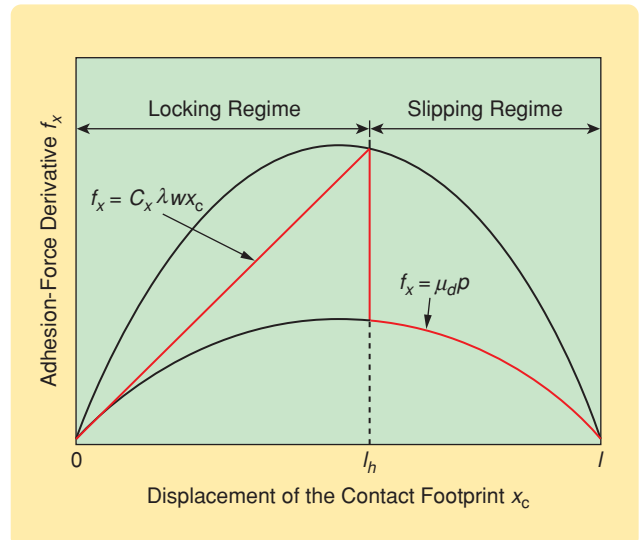
FIGURE 6 A typical distribution of the tangential force coefficient in a contact footprint. In the locking regime, the adhesion-force derivative f_x increases linearly according to the displacement of the contact footprint. However, in the slipping regime, the adhesion-force derivative f_x is proportional to the contact pressure. The adhesion force F_a can be calculated by integrating the adhesion-force derivative f_x over the length of the contact footprint.

where z is the average bristle deflection, v_s is the Stribeck velocity, and F_s is the static friction force. In addition, α, α_1, and α_2 are the bristle stiffness coefficient, bristle damping coefficient, and viscous damping coefficient, respectively.

The functions $g(\sigma)$ and F in (14) and (15) are determined by selecting the exponential term in (14) and coefficients α, α_1, and α_2 in (15), respectively, to match the mathematical model with the measured friction. For example, to match the mathematical model with the measured friction, the standard LuGre model is modified by using $e^{-|\sigma/v_s|^{1/2}}$ in place of the term $e^{-(\sigma/v_s)^2}$ in (14). Furthermore, for the tire model for vehicle traction control, the function F given by (15) is modified by including the normal force. Thus, (13)–(15) are modified as [26]

FIGURE 7 Typical shape of the general friction force and adhesion force in railway rolling stocks according to the relative velocity. Static friction, Coulomb friction, the Stribeck effect, and viscous friction are included in the general friction force, while static friction, Coulomb friction, and the Stribeck effect are included in the adhesion force in railway rolling stocks.
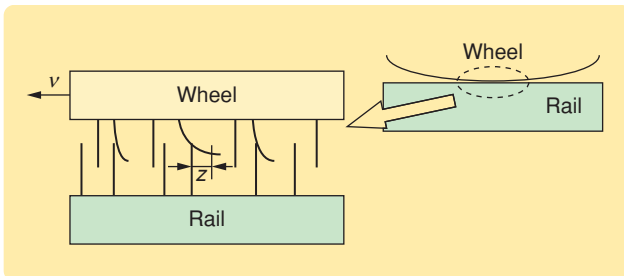
FIGURE 8 Bristle model between the rail and the wheel. The friction interface between the rail and the wheel is modeled as contact between bristles. For simplicity the bristles on the lower part are shown as being rigid. [11].

$$\frac{dz}{dt} = \sigma - \frac{\alpha'|\sigma|}{g(\sigma)}z, \tag{16}$$

$$g(\sigma) = \mu_c + (\mu_s - \mu_c)e^{-|\sigma/v_s|^{\frac{1}{2}}}, \tag{17}$$

$$F = (\alpha'z + \alpha_1'\dot{z} + \alpha_2'\sigma)N, \tag{18}$$

where μ_s and μ_c are the static friction coefficient and Coulomb friction coefficient, respectively, $N = mg$ is the normal force, m is the mass of the wheel, and $\alpha' = \alpha/N$, $\alpha_1' = \alpha_1/N$, and $\alpha_2' = \alpha_2/N$ are the normalized wheel longitudinal lumped stiffness coefficient, normalized wheel longitudinal lumped damping coefficient, and normalized viscous damping coefficient, respectively.

In general, it is difficult to measure and identify all six parameters, α, α_1, α_2, F_s, F_c, and v_s in the LuGre model equations. In particular, identifying friction coefficients such as α and α_1 requires a substantial amount of experimental data [27]. We thus develop a dynamic model for friction phenomena in railway rolling stocks, as shown in Figure 7. The dynamic model retains the simplicity of the Dahl model while capturing the Stribeck effect.

As shown in Figure 8 [11], the motion of the bristles is assumed to be the stress-strain behavior in solid mechanics, which is expressed as

$$\frac{dF_a}{dx} = \alpha\left[1 - h(\sigma)F_a\right], \tag{19}$$

where F_a is the adhesion force, α is the coefficient of the dynamic adhesion force, and x and σ are the relative displacement and velocity of the contact surface, respectively. In addition, the function $h(\sigma)$ is selected according to the friction characteristics.

Defining z to be the average deflection of the bristles, the adhesion force F_a is assumed to be given by

$$F_a = \alpha z. \tag{20}$$

The derivative of F_a can then be expressed as

$$\frac{dF_a}{dt} = \frac{dF_a}{dx}\frac{dx}{dt} = \frac{dF_a}{dx}\sigma = \alpha[1 - h(\sigma)F_a]\sigma = \alpha\frac{dz}{dt}. \tag{21}$$

It follows from (20) and (21) that the internal state z is given by

$$\dot{z} = \sigma[1 - h(\sigma)F_a] = \sigma[1 - \alpha h(\sigma)z]. \tag{22}$$

To select the function $h(\sigma)$ for railway rolling stocks, the term $e^{-\sigma/v_s}$ is used in place of $e^{-(\sigma/v_s)^2}$ in (14). This term is
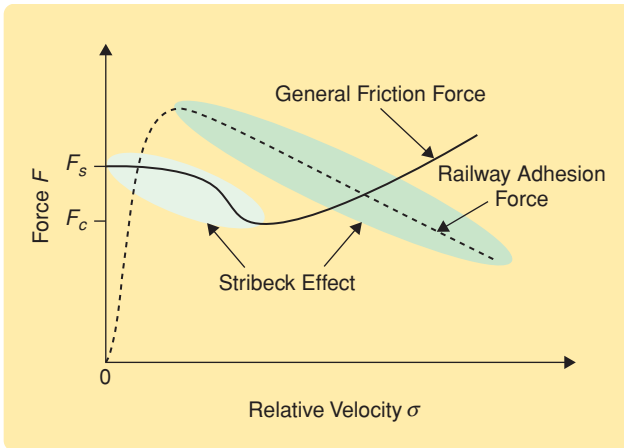
simplified by executing the Taylor series expansion for $e^{-\sigma/v_s}$ and by taking only the linear term $1 - \sigma/v_s$. In addition, neglecting the coefficients α_1 and α_2 in (15) for simplicity yields

$$g(\sigma) = F_c + (F_s - F_c)\left(1 - \frac{\sigma}{v_s}\right) = F_s - (F_s - F_c)\frac{\sigma}{v_s}. \quad (23)$$

By comparing (13) and (23) with (22) and by considering the relative velocity σ, which is positive in railway rolling stocks, $h(\sigma)$ in (22) can be derived as

$$h(\sigma) = \frac{\beta}{\gamma - \sigma}, \quad (24)$$

where $\beta = (1/F_s - F_c)v_s$ and $\gamma = (F_s/F_s - F_c)v_s$. In general, β and γ are positive tuning parameters because F_s is larger than F_c as shown in Figure 7. In the dynamic model, the parameter α is the coefficient for the starting point of the slip regime, where the adhesion force decreases according to the relative velocity, and the parameters β and γ are the coefficients for the slope and shift in the slip regime, respectively.

VERIFICATION OF THE ADHESION FORCE MODELS

To verify the adhesion force models, experiments using a braking performance test rig in the Railway Technical Research Institute in Japan and computer simulations are carried out under various initial braking velocity conditions. Figure 9 shows the test rig for the braking performance test. The conceptual schematic diagram is shown in Figure 10. This test rig consists of a main principal axle with a wheel for rolling stocks on a rail, flywheels, a main motor, a sub-axle with a wheel, and a brake disk. After accelerating to the target velocity by the main motor, the brake caliper applies a brake force to the wheel. The inertia of the flywheels plays the role of the inertia of the running railway rolling stocks.

The test conditions are shown in Table 1. During the experiments, the brake torque T_b, the wheel load N, the angular velocity of the wheel ω, and the velocity of the rolling stocks v are measured simultaneously. The adhesion torque T_a between the rail and the wheel used in the calculation of the adhesion coefficient is also estimated in real time. As in the case of running vehicles, it is impossible to measure the adhesion torque directly on the brake performance test rig.

It is essential that knowledge of the adhesion torque be available for both ABS in automobiles and wheel-slip control of rolling stocks. However, it is difficult to directly acquire this information. While an optical sensor, which is expensive [14], can be used to acquire this information, the adhesion force between the wheel and the rail is estimated through the application of a Kalman filter [28]. By using this scheme, the adhesion force can be estimated online during the normal running of the vehicle before the brake is
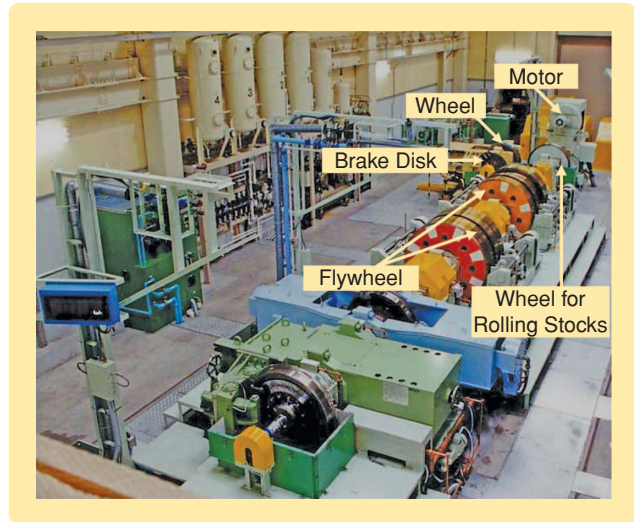
FIGURE 9 Test rig for the brake performance test. This test rig consists of a main principal axle with a wheel for rolling stocks on a rail, flywheels, a main motor, a subaxle with a wheel, and a brake disk.
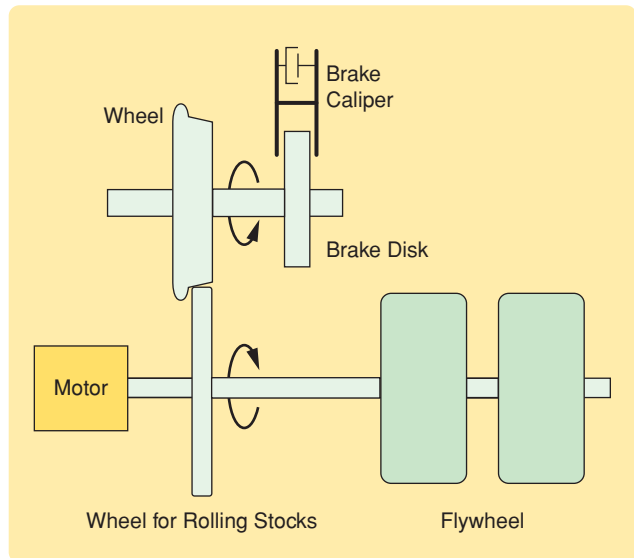
FIGURE 10 Conceptual schematic diagram of the test rig for the brake performance test. After accelerating to the target velocity by the main motor, the brake caliper applies a brake force to the wheel. The inertia of the flywheels plays the role of the inertia of the running railway rolling stocks.

TABLE 1 Test conditions of the test rig for the brake performance test. During the experiments, the brake torque T_b, the wheel load N, the wheel angular velocity ω, and the velocity v of the rolling stocks are measured simultaneously.

Test Condition	Value
Initial braking velocity	30, 60, 100, 140 km/h
Slip ratio	0–50%
Wheel load	34.5 kN
Wheel inertia	60.35 kg-m^2
Viscous friction torque coefficient	0.25 N-m-s

applied. A disturbance observer considering the first resonant frequency of the rolling stocks is designed to avoid undesirably large wheel slip, which causes damage to the rail and wheel [29]. A sliding mode adhesion-force observer using the estimation error of the wheel angular velocity and based on a LuGre model can be used for this purpose [30].

We now consider an adhesion-torque observer for estimation. In (3), we neglect the unknown disturbance torque of the wheel T_d because the dominant disturbance torque caused by the vibration of the brake caliper acts only for a moment in the initial braking time. Then the adhesion torque T_a is expressed as

$$T_a = J\dot{\omega} + B\omega + T_b. \tag{25}$$

FIGURE 11 Adhesion-torque observer. This observer serves as a disturbance observer. A first-order lowpass filter is used to implement the differential term, which can be removed by an equivalent transformation of the block diagram. The estimated adhe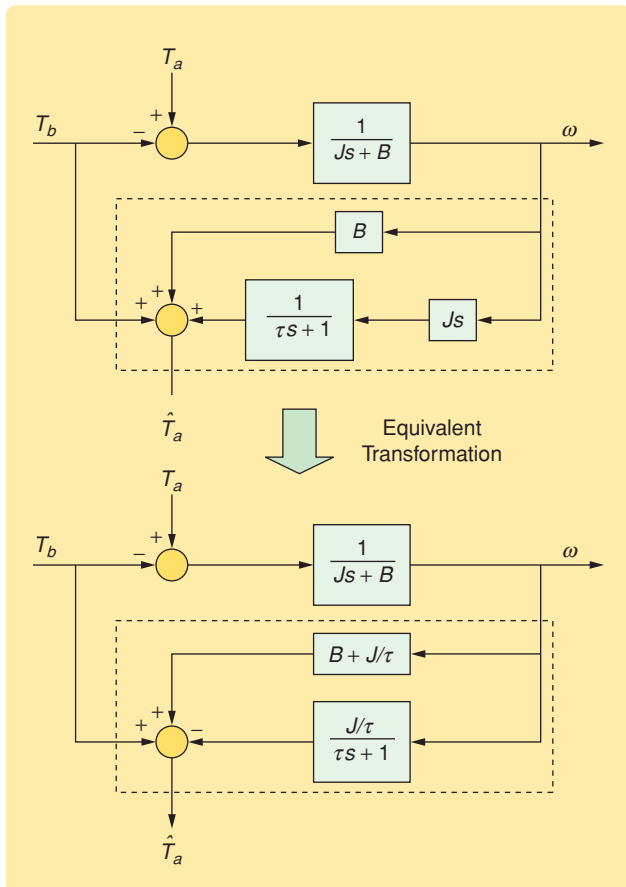sion torque is obtained by the brake torque and the angular velocity of the wheel through the first-order lowpass filter.

Taking Laplace transforms yields

$$T_a(s) = Js\omega(s) + B\omega(s) + T_b(s). \tag{26}$$

Since a differential term is included in (26), we implement a first-order lowpass filter of the form

$$\hat{T}_a(s) = \frac{Js}{\tau s + 1}\omega(s) + B\omega(s) + T_b(s), \tag{27}$$

or

$$\hat{T}_a(s) = \left(B + \frac{J}{\tau} - \frac{J/\tau}{\tau s + 1}\right)\omega(s) + T_b(s), \tag{28}$$

where τ is the time constant of the lowpass filter in the adhesion-torque observer, which is illustrated in Figure 11. The estimated adhesion coefficient $\hat{\mu}$ can now be obtained by

$$\hat{\mu} = \frac{\hat{T}_a}{Nr}. \tag{29}$$

As shown by the experimental wheel-slip results in Figure 12, before 4.5 s, the velocity v of the rolling stocks matches the tangential velocity $v_w = r\omega$ of the wheel, where r and ω are the radius and angular velocity of the wheel, respectively, while a large difference occurs
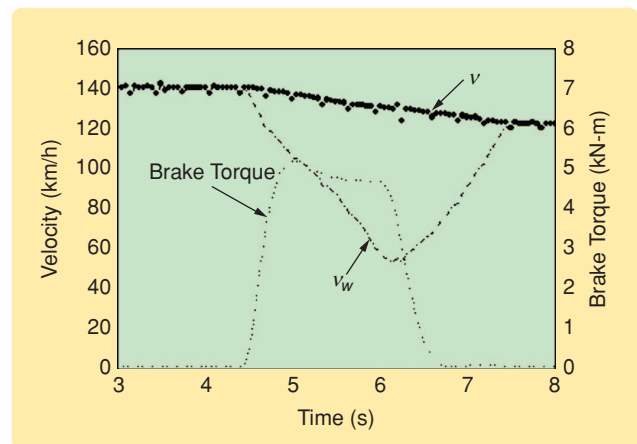
FIGURE 12 Experimental wheel-slip results. Before 4.5 s, the velocity v of the rolling stocks matches the tangential velocity $v_w = r\omega$ of the wheel, where r and ω are the radius and angular velocity of the wheel, respectively, while a large difference occurs between the velocity of the rolling stocks and the tangential velocity of the wheel at 4.5 s when a large brake torque is applied. This difference means that a large wheel slip occurs as a result of braking. The controller ceases the braking action at 6.1 s when the slip ratio exceeds 50%.

between the velocity of the rolling stocks and the tangential velocity of the wheel at 4.5 s when a large brake torque is applied. This difference means that large wheel slip occurs as a result of braking. The controller ceases the braking action at 6.1 s when the slip ratio exceeds 50%. Henceforth, the tangential velocity of the wheel recovers, and the slip ratio decreases to zero by the adhesion force between the rail and the wheel. In the experiment, to prevent damage due to excessive wheel slip, the applied brake torque is limited so that the slip ratio does not exceed 50%.

Table 2 shows the parameters of the adhesion force models for computer simulation. In Table 2, the parameter values for the length l and the width w of the contact footprint are taken from [31]. The constant a in (7) for the

TABLE 2 Parameters of the beam and bristle models for computer simulation. The parameter values for the length l and width w of the contact footprint are taken from [31]. The velocity v_0 is the initial braking velocity of the rolling stocks.

Parameter	Notation	Value
Modulus of transverse elasticity	C_x	1.52×10^9 N/m^2
Length	l	0.019 m
Width	w	0.019 m
Wheel load	N	34.5 kN
Maximum adhesion coefficient for v_0 = 30, 60, 100, 140 km/h	μ_{max}	0.360, 0.310, 0.261, 0.226
Radius of the wheel	r	0.43 m

FIGURE 13 Experimental and simulation results of the adhesion coefficient according to the slip ratio and initial braking velocity. The symbols "△", "○", and "×" represent the maximum, mean, and minimum values of the adhesion coefficient, respectively, which are obtained by experiments. Compared with the experimental results, the bristle model describes the adhesion coefficient more accurately according to the slip ratio and initial braking than the beam model. (a) Initial braking velocity $v_0 = 140$ km/h. (b) Initial braking velocity $v_0 = 100$ km/h. (c) Initial braking velocity $v_0 = 60$ km/h. (d) Initial braking velocity $v_0 = 30$ km/h.

TABLE 3 Mean values of the absolute errors between the experimental results for the mean value of the adhesion coefficient and the simulation results for the beam and bristle models according to the initial braking velocity of the rolling stocks. The mean values of the absolute errors in the relevant range of the initial braking velocity for the beam and bristle models are 0.011 and 0.0083 km/h, respectively. Using the bristle model in place of the beam model yields 24.5% improvement in accuracy.

	Initial braking velocity			
Adhesion model	30 km/h	60 km/h	100 km/h	140 km/h
Beam model	0.0130	0.0085	0.0132	0.0093
Bristle model	0.0080	0.0080	0.0102	0.0077

beam model is determined as 0.0013 h/km based on the adhesion experimental results at the initial braking velocity of 140 km/h.
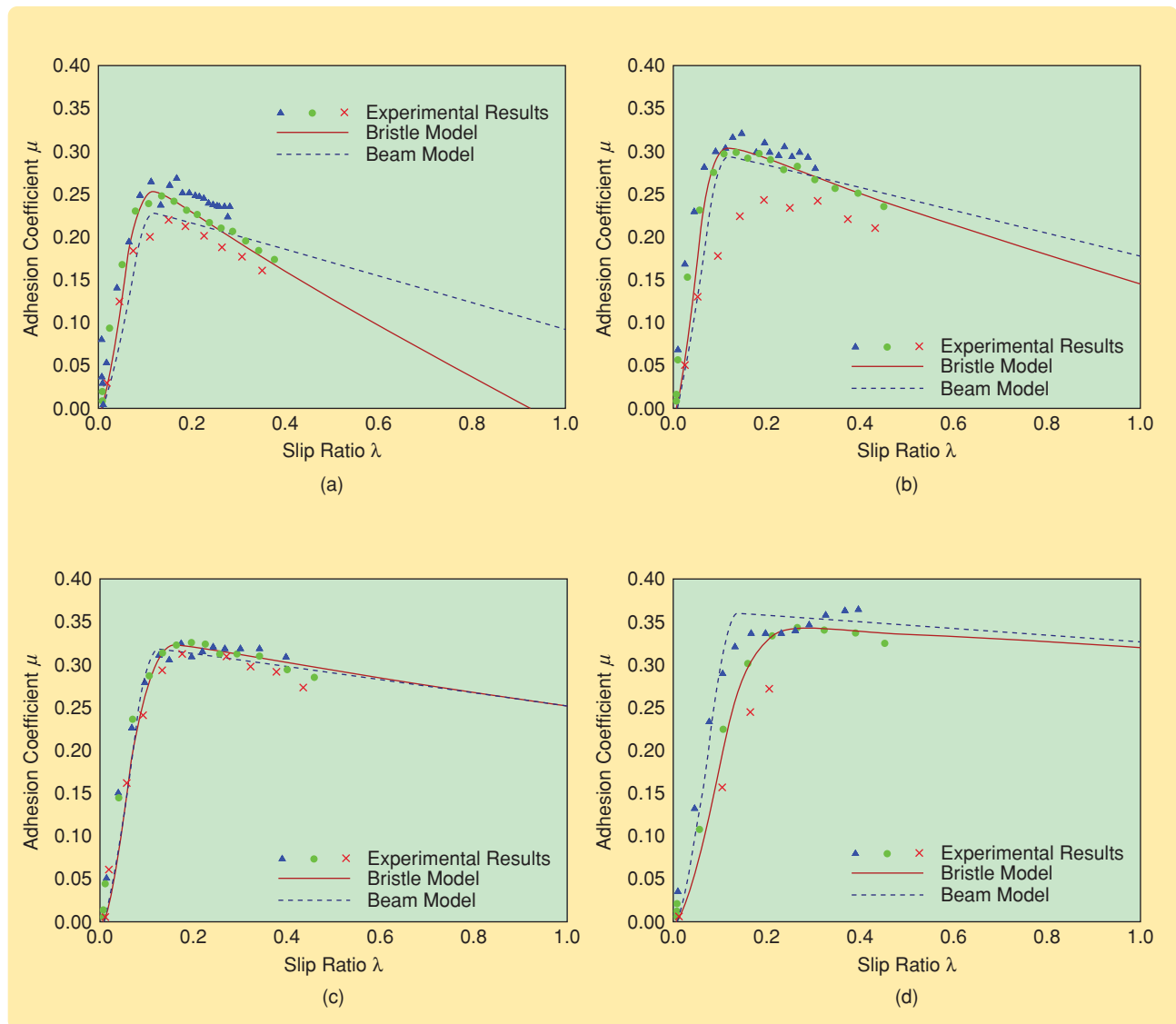
Figure 13 shows experimental and simulation results of the adhesion coefficient according to the slip ratio and initial braking velocity. As shown in Figure 13, the variation of the adhesion coefficients obtained by the experiments is large. It is therefore difficult to determine a precise mathematical model for the adhesion force. In spite of these large variations, it is found that the experimental results of the mean value of the adhesion coefficient according to the slip ratio are consistent with the simulation results based on the two kinds of adhesion force models. Table 3 shows the mean values of the absolute errors between the experimental results for the mean value of the adhesion coefficient and the simulation results for the beam and bristle models according to the initial braking velocity of the rolling stocks. Mean values of the absolute errors in the relevant range of the initial braking velocity for the beam and bristle models are 0.011 and 0.0083, respectively. Using the bristle model in place of the beam model yields a 24.5% improvement in accuracy.

From the experimental results in Figure 13, the parameters α, β, and γ of the bristle model (19)–(22) and (24) can be expressed as

$$\alpha = 5.455 \times 10^4 - 3.641 \times 10^2 v_0 + 3.798 \times 10^{-1} v_0^2, \quad (30)$$
$$\beta = 1.873 \times 10^{-2} - 6.059 \times 10^{-5} v_0 + 5.500 \times 10^{-8} v_0^2, \quad (31)$$
$$\gamma = 2.345 \times 10^2 - 8.620 \times 10^{-1} v_0 + 1.053 \times 10^{-4} v_0^2, \quad (32)$$

where v_0 is the initial braking velocity of the rolling stocks. The coefficients in (30)–(32) are obtained by curve fitting for the values of the parameters according to the initial braking velocity.

Simulation results of the mean value of the adhesion coefficients for the beam model and bristle model according to the slip ratio and initial braking velocity, respectively, are shown in figures 14 and 15. These results show a similar tendency for the change in the initial braking velocity conditions. However, the adhesion force model based on the beam model cannot represent the dynamic characteristics of friction. The beam model is obtained by curve fitting the experimental results on the adhesion force, while the bristle model, which includes the friction dynamics, describes the effect of the initial braking velocity accurately in the adhesion regime, where the adhesion
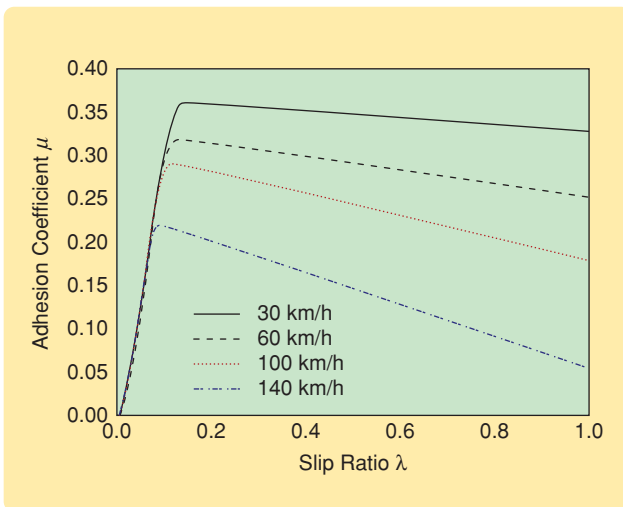
FIGURE 14 Simulation results of the mean value of the adhesion coefficient for the beam model according to the slip ratio and initial breaking velocity (30, 60, 100, 140 km/h). The mean value of the absolute errors between the adhesion coefficients based on the beam model and experiment in Figure 13 is 0.0193 in the adhesion regime, where the adhesion coefficient increases according to the slip ratio.
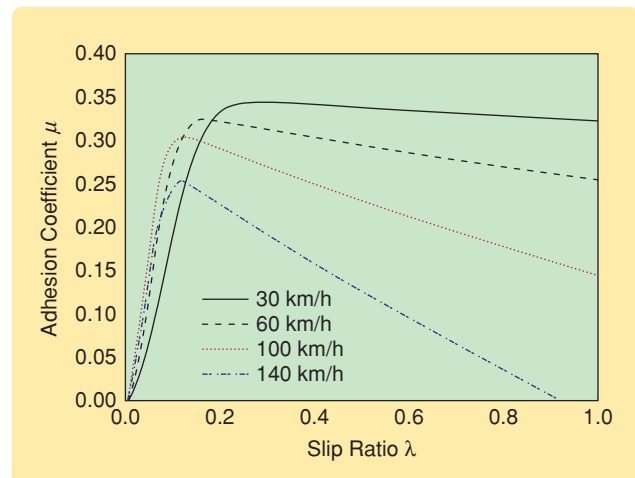
FIGURE 15 Simulation results of the mean value of the adhesion coefficient for the bristle model according to the slip ratio and initial braking velocity. The mean value of the absolute errors between the adhesion coefficients by the bristle model and experiment in Figure 13 is 0.0138 in the adhesion regime. Therefore, in the adhesion regime, the bristle model describes the effect of the initial braking velocity more accurately than the beam model.

force increases according to the slip ratio, as shown in Figure 15. Therefore, the bristle model is more applicable than the beam model for the desired wheel-slip controller design.

DESIRED WHEEL SLIP USING ADAPTIVE SLIDING MODE CONTROL

The desired wheel-slip brake control system is designed by using an ASMC scheme to achieve robust wheel-slip brake control. In the controller design process, the random value of adhesion torque, the disturbance torque due to the vibration of the brake caliper, and the traveling resistance force of the rolling stocks are considered as system uncertainties. The mass of the rolling stocks and the viscous friction torque coefficient are also considered as parameters with unknown variations. The adaptive law for the unknown parameters is based on Lyapunov stability theory.

The sliding surface s for the design of the adaptive sliding mode wheel-slip brake control system is defined as

$$s = e + \rho \int_0^t e\,dt, \tag{33}$$

where $e = \sigma_d - \sigma$ is the tracking error of the relative velocity, $\sigma = \lambda v = v - r\omega$ is the relative velocity, σ_d is the reference relative velocity, and ρ is a positive design parameter.

The sliding mode control law consists of equivalent and robust control terms, that is,

$$T_b = U_{eq} + U_r, \tag{34}$$

where U_{eq} and U_r are the equivalent and robust control terms. To obtain U_{eq} and U_r, we combine (3), (4) with the derivative of the sliding surface in (33) and include random terms in the adhesion force $F_{ar} = F_a + F_r$ and the adhesion torque $T_{ar} = T_a + T_r$, where F_r and T_r are the random terms of the adhesion force and adhesion torque, respectively. Then, the derivative of the sliding surface can be written as

$$\dot{s} = \dot{\sigma}_d + \left(\frac{1}{rM} + \frac{r}{J}\right)T_a + \left(\frac{1}{rM} + \frac{r}{J}\right)T_r + \frac{1}{M}F_r$$

$$- \frac{r}{J}T_b - \frac{rB}{J}\omega - \frac{r}{J}T_d + \rho\,e. \tag{35}$$

To determine the equivalent control term U_{eq}, uncertainties such as random terms in the adhesion force and adhesion torque F_r and T_r, as well as the disturbance torque T_d in (35) are neglected, and it is assumed that the

sliding surface s is at steady state, that is, $\dot{s} = 0$, then the equivalent control law can be determined as

$$U_{eq} = \frac{J}{r}\left[\dot{\sigma}_d + \left(\frac{1}{rM} + \frac{r}{J}\right)T_a - \frac{rB}{J}\omega + \rho e\right]. \tag{36}$$

Thus, $\dot{s}$ can be rewritten as

$$\dot{s} = \left(\frac{1}{rM} + \frac{r}{J}\right)T_r + \frac{1}{M}F_r - \frac{r}{J}T_d - \frac{r}{J}U_r. \tag{37}$$

In the standard sliding mode control, to satisfy the reachability condition that directs system trajectories toward a sliding surface where they remain, the derivative of the sliding surface is selected as

$$\dot{s} = -K\,\mathrm{sgn}(s). \tag{38}$$

In this case, chattering occurs in the control input. To attenuate chattering in the control input, the derivative of the sliding surface is selected as [32], [33]

$$\dot{s} = -Ds - K\,\mathrm{sgn}(s), \tag{39}$$

where the parameters D and K are positive.

To determine a control term U_r that achieves robustness to uncertainties such as random terms in the adhesion force and adhesion torque, as well as the disturbance torque, it is assumed that

$$D_0|s| + K_0 > \left|\left(\frac{1}{rM} + \frac{r}{J}\right)T_r + \frac{1}{M}F_r - \frac{r}{J}T_d\right| + \eta, \tag{40}$$

where the parameters $D_0 = (r/J)D$, $K_0 = (r/J)K$, and η are positive. Then, the robust control law can be determined as

$$U_r = Ds + K\,\mathrm{sgn}(s), \tag{41}$$

and using (40), the reachability condition is satisfied as

$$s\dot{s} = s\left[\left(\frac{1}{rM} + \frac{r}{J}\right)T_r + \frac{1}{M}F_r - \frac{r}{J}T_d - D_0 s - K_0\,\mathrm{sgn}(s)\right]$$

$$\leq |s|\left[\left|\left(\frac{1}{rM} + \frac{r}{J}\right)T_r + \frac{1}{M}F_r - \frac{r}{J}T_d\right| - D_0|s| - K_0\right]$$

$$< -\eta|s|. \tag{42}$$

Finally, the sliding mode control law is selected as

$$T_b = U_{eq} + U_r$$

$$= \frac{J}{r}\left[\dot{\sigma}_d + \left(\frac{1}{rM} + \frac{r}{J}\right)T_a - \frac{rB}{J}\omega + \rho e\right] + Ds + K\operatorname{sgn}(s),$$

(43)

where the reference slip acceleration $\dot{\sigma}_d$ and the adhesion torque T_a cannot be measured during operation. Therefore, to implement the control system, the reference slip acceleration $\dot{\sigma}_d = \lambda_d \dot{v}$ must be estimated by $\lambda_d \hat{v}$, where $\hat{v}$ is the estimated acceleration of the rolling stocks, which can be obtained by the measured velocity of the rolling stocks through the first-order filter $G_f(s) = s/(\tau s + 1)$. In addition, the adhesion torque $T_a = rF_a$ must be replaced by the calculated value given by (20) and (22) with the measured relative velocity σ.

If the mass of the rolling stocks M and the viscous friction torque coefficient B are considered as parameters with variation, that is, $M = M_n + M_p$ and $B = B_n + B_p$, where the subscripts n and p denote the nominal and perturbation values, respectively, then the uncertainty ψ in the mass of the rolling stocks and the viscous friction torque coefficient is defined as

$$\psi = \frac{1}{rM_p}T_m - \frac{rB_p}{J}\omega = \theta^T \varphi,$$

(44)

where

$$\theta^T = \left[\begin{array}{cc} \dfrac{1}{rM_p} & -\dfrac{rB_p}{J} \end{array}\right], \qquad \varphi = \left[\begin{array}{c} T_m \\ \omega \end{array}\right].$$
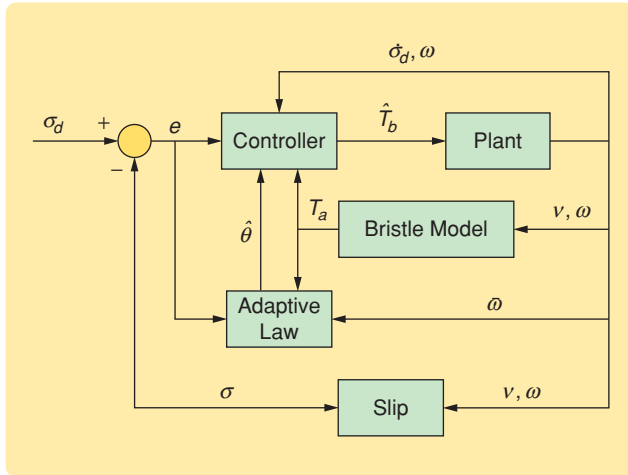
FIGURE 16 Wheel-slip control system. The reference relative velocity $\sigma_d = \lambda_d v$ is given by the constant reference slip ratio, and the reference relative acceleration $\dot{\sigma}_d$ cannot be measured in real time. Therefore, to implement the control system, the relative acceleration $\dot{\sigma}_d = \lambda_d \dot{v}$ must be estimated by $\lambda_d \hat{v}$, where $\hat{v}$ is the estimated acceleration of the rolling stocks, which can be obtained from the measured velocity of the rolling stocks through the first-order filter $G_f(s) = s/(\tau s + 1)$.

The parameter vector θ is considered as an unknown parameter vector, which can be estimated by using the update law. From (43) and the estimated unknown parameter vector $\hat{\theta}$, the estimated sliding model control law can be selected as

$$\hat{T}_b = \frac{J}{r}\left[\dot{\sigma}_d + \left(\frac{1}{rM_n} + \frac{r}{J}\right)T_m - \frac{rB_n}{J}\omega + \hat{\theta}^T\varphi + \rho e\right]$$

$$+ Ds + K\operatorname{sgn}(s).$$

(45)

To obtain the update law for the unknown parameters, we consider the Lyapunov candidate

$$V = \frac{1}{2}s^2 + \frac{1}{2k}\tilde{\theta}^T\tilde{\theta},$$

(46)

where $\tilde{\theta} = \theta - \hat{\theta}$, θ and $\hat{\theta}$ are the nominal and estimated parameter vectors, respectively, and k is a positive parameter. The derivative of the Lyapunov candidate including sliding dynamics is expressed as

$$\dot{V} = s\left[\dot{\sigma}_d + \left(\frac{1}{rM_n} + \frac{r}{J}\right)T_m - \frac{rB_n}{J}\omega - \frac{r}{J}\hat{T}_b + \theta^T\varphi + \rho e\right]$$

$$- \frac{1}{k}\tilde{\theta}^T\dot{\hat{\theta}}.$$

(47)

Substituting the estimated brake torque $\hat{T}_b$ given by (46) into (47) yields

$$\dot{V} = -Ds^2 - Ks\operatorname{sgn}(s) + \tilde{\theta}^T\left(s\varphi - \frac{1}{k}\dot{\hat{\theta}}\right).$$

(48)

By using the update law for the unknown parameters given by

$$\dot{\hat{\theta}} = ks\varphi,$$

(49)

the derivative of the Lyapunov candidate (48) is nonpositive. The invariant set theorem then guarantees asymptotic stability of the wheel-slip brake control system [34].

PERFORMANCE EVALUATION OF THE DESIRED WHEEL-SLIP CONTROL SYSTEM

The characteristics of the wheel-slip control system shown in Figure 16 are evaluated by simulation. The performance and robustness of the wheel-slip control system using the ASMC scheme are evaluated for railway rolling stocks, while considering system uncertainties such as parameter variation, railway conditions, disturbances, and unmodeled dynamics.

For simulation, the bristle model is used for the adhesion force model because the bristle model is relatively close to the actual adhesion force compared with the beam model. In addition, it is assumed that the brake torque is applied when the velocity of the rolling stocks is 100 km/h. From the experimental results in Figure 13,

Control scheme	Control gain	Beam model	Bristle model
PI	K_p	400 N-h	650 N-h
	K_i	54 N	27 N
ASMC	D	1.62 h^{-1}	1.53 h^{-1}
	K	70 km/h^2	70 km/h^2
	ρ	1.65 h^{-1}	1.54 h^{-1}
	k	2.1×10^{-8}	2.1×10^{-8}

it is assumed that the random adhesion force F_r is a white noise signal with a Gaussian distribution that has a standard deviation of 0.431 kN. Since the actual brake force is applied to the wheel disk by the brake caliper, the vibration occurs on the brake caliper at the initial braking moment. Therefore, the disturbance torque $T_d = 0.05 T_b e^{-4t} \sin 10\pi t$, caused by the vibration of the brake caliper, is considered in the simulation. In addition, the traveling resistance force $F_r = 0.63 v^2$ of the rolling stocks and the viscous friction $B = 0.25^{-0.01t}$ is considered, which causes overheating between the wheel disk and the brake pad. Finally, the unmodeled dynamics $G_a(s) = e^{-0.15s}/(0.6s + 1)$ of the pneumatic actuator of the brake control system are considered.

To assess the braking performance in the presence of parameter variations, the simulation is carried out under the assumption that the mass of the rolling stocks changes according to the number of passengers and that the rolling stocks travel in dry or wet rail conditions. It is assumed that the mass of the rolling stocks changes from 3517 to 5276 kg at 25 s. It is also assumed that the maximum adhesion force under wet rail conditions is approximately half of the maximum adhesion force under dry rail conditions and that the rail conditions change from dry to wet at 25 s. Figure 17 shows the relationship between the adhesion force coefficient and the slip ratio according to the change in rail conditions from dry to wet based on the beam and bristle models, which are considered in the simulation. As shown in Figure 17, the reference slip ratio is assumed to be 0.119 and 0.059 under dry and wet rail conditions, respectively, for the beam model, and 0.132 and 0.092 under dry and wet rail conditions, respectively, for the bristle model.

To verify the performance and robustness of the ASMC system, the desired wheel-slip control system using the ASMC scheme is compared with a PI control system through simulation. Control gains of the PI and ASMC systems are selected through trial and error by considering various constraints for each case, such as the maximum brake torque and the maximum slip ratio

allowed until the desired performance and robustness are obtained, which are summarized in Table 4. In controller design, the bristle model and beam model are considered for the adhesion force model. Figures 18 and 19 show the velocities of the wheel and rolling stocks for the PI and ASMC systems based on the beam and bristle models, respectively. As shown in Figure 18, the braking distance and time of the PI control system until the velocity of the rolling stocks reaches 5 km/h are 700 m and 59.5 s, respectively, for the PI control based on the beam model, and 682 m and 58.9 s, respectively, for the PI control based on the bristle model. By using the PI control based on the bristle model in place of the PI control based on the beam model, the braking distance and time are improved by 2.6% and 1%, respectively. However, the PI control system cannot effectively compensate for system uncertainties such as the mass of the
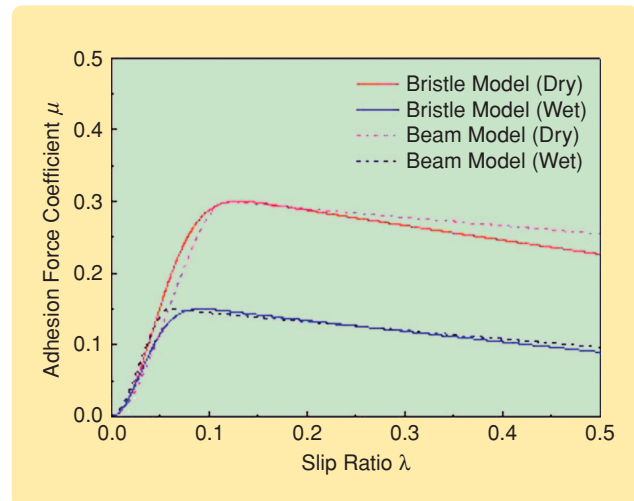
FIGURE 17 Relationship between the adhesion force coefficient and the slip ratio according to the change in rail conditions from dry to wet based on the beam and bristle models. For the beam model, the reference slip ratio is assumed to be 0.119 and 0.059 under dry and wet rail conditions, respectively, whereas, for the bristle model, the reference slip ratio is assumed to be 0.132 and 0.092 under dry and wet rail conditions, respectively.

rolling stocks, railway conditions, the traveling resistance force, and variations of the viscous friction coefficient.

As shown in Figure 19 for the ASMC system, the braking distance and time are 607 m and 55.3 s, respectively, for the ASMC based on the beam model and 581 m and 50.7 s, respectively, for the ASMC based on the bristle model. Figure 19 shows that the ASMC system provides robust velocity regulation of the rolling stocks in the presence of variations in the mass of the rolling stocks and rail conditions. In this case, the braking distance and time are improved by 4.3% and 8.3%, respectively, by using the ASMC based on the bristle model in place of the ASMC based on the beam model.

Figure 20 shows the brake torques for the PI and ASMC systems based on the beam and bristle models. The expended braking energies of the PI and ASMC systems during braking time are 1.77×10^7 N-m and 1.71×10^7 N-m, respectively. Therefore, by using the ASMC system, it is possible to effectively reduce the braking time and distance using a relatively small braking energy consumption.

The operation of the PI and ASMC wheel-slip control systems can also be demonstrated through the slip ratios. Figure 21 shows the slip ratios of the PI and ASMC sys-

tems based on the beam and bristle models. Figure 21 shows that the PI control system has a large tracking error of slip ratio compared with the ASMC system. However, the wheel-slip control system using the ASMC scheme can maintain the slip ratio near the reference slip ratio during the braking time although the slip ratios fluctuate slightly after 25 s when the system uncertainties are applied. Therefore, it is appropriate to use the ASMC system to obtain the maximum adhesion force and a short braking distance. Using the ASMC based on the bristle model in place of the PI control based on the beam model yields 28% improvement in the wheel slip. Table 5 summarizes the performance of the PI and ASMC systems based on the beam and bristle models.

CONCLUSIONS

Two types of models, namely, the beam and bristle models, for the adhesion force in railway rolling stocks are developed. The validity of the beam and bristle models is determined through an adhesion test using a brake performance test rig. By comparing the simulation results of the two types of adhesion force models with the experimental results, it is found that these adhesion
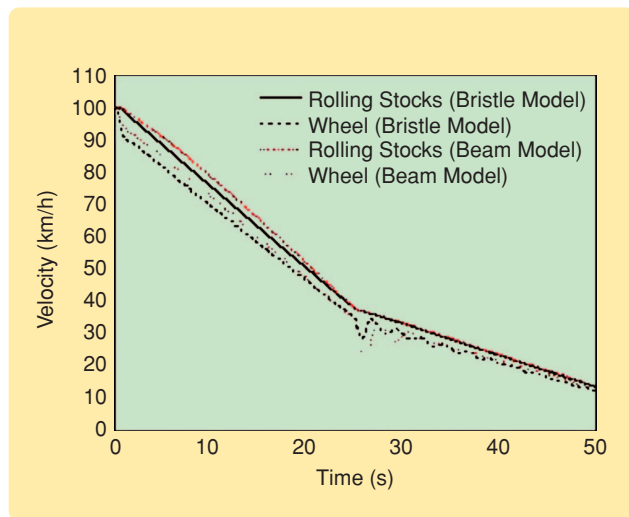
FIGURE 18 Velocities of the wheel and rolling stocks for the proportional-integral (PI) control systems based on the beam and bristle models. For the PI control based on the beam model, the braking distance and time of the PI control system until the velocity of the rolling stocks reaches 5 km/h are 700 m and 59.5 s, respectively, and 682 m and 58.9 s, respectively, for the PI control based on the bristle model. By using the PI control based on the bristle model in place of the PI control based on the beam model, the braking distance and time are improved by 2.6% and 1%, respectively.
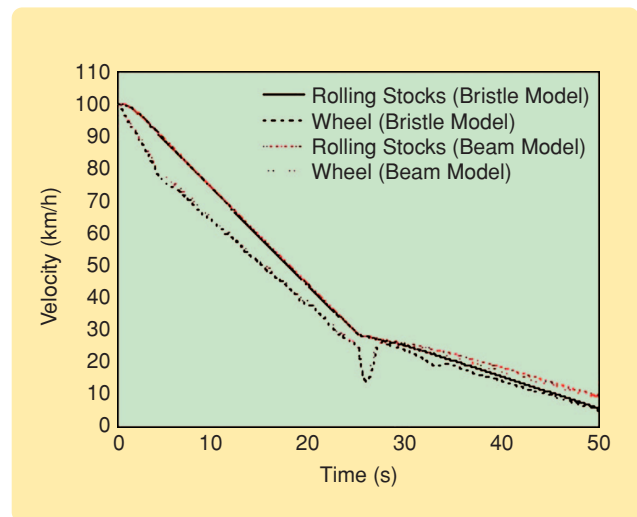
FIGURE 19 Velocities of the wheel and rolling stocks for the adaptive sliding mode control (ASMC) systems based on the beam and bristle models. For the ASMC based on the beam model, the braking distance and time of the ASMC system until the velocity of the rolling stocks reaches 5 km/h are 607 m and 55.3 s, respectively, and 581 m and 50.7 s, respectively, for the ASMC based on the bristle model. By using the ASMC based on the bristle model in place of the ASMC based on the beam model, the braking distance and time are improved by 4.3% and 8.3%, respectively.

force models can effectively represent the experimental results. However, the adhesion force model based on the beam model cannot represent the dynamic characteristics of friction, while the bristle model can mathematically include the dynamics on friction and can precisely consider the effect of the initial braking velocity in the adhesion regime. Therefore, the bristle model is more appropriate than the beam model for the design of the wheel-slip controller.

In addition, based on the beam and bristle models, the PI and ASMC systems are designed to control wheel slip in railway rolling stocks. Through simulation, we evaluate the performance and robustness of the PI and ASMC systems based on the beam and bristle models for railway rolling stocks. It is verified from the simulation study that, among the four types investigated according to control schemes and adhesion force models, the ASMC system based on the bristle model is the most suitable system for the wheel slip in the rolling stocks with system uncertainties such as the mass and traveling resistance force of the rolling stocks, rail conditions, random adhesion torque, disturbance torque due to the vibration of the brake caliper, and unmodeled actuator dynamics.
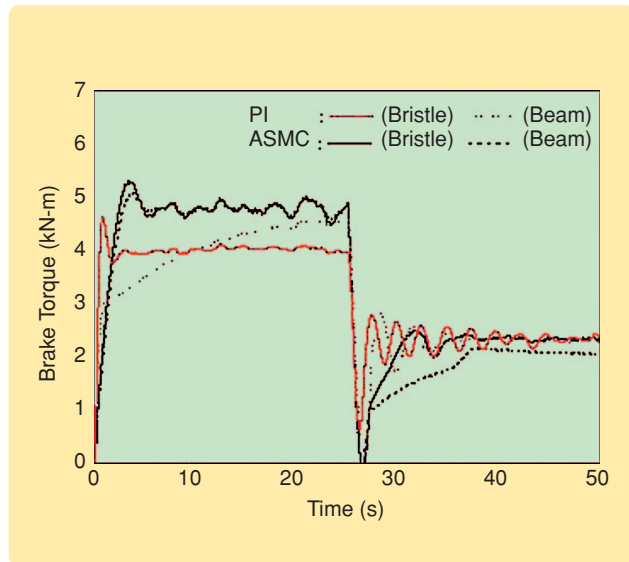

FIGURE 20 Brake torques for the proportional-integral (PI) and adaptive sliding mode control (ASMC) systems based on the beam and bristle models. The expended braking energies of the PI and ASMC systems during braking time are 1.77×10^7 N-m and 1.71×10^7 N-m, respectively.
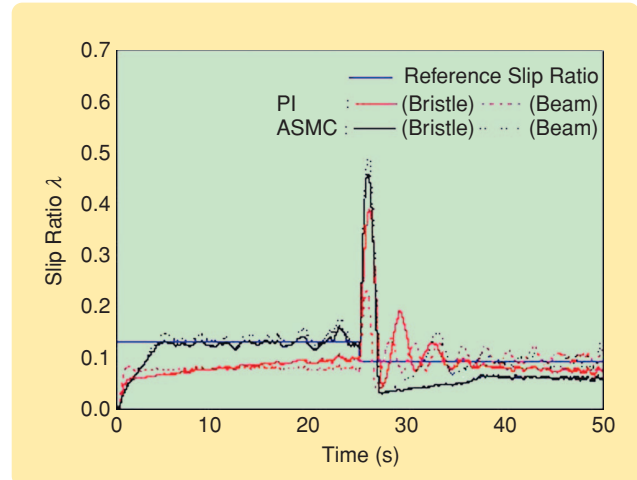
FIGURE 21 Slip ratios of the proportional-integral (PI) and adaptive sliding mode control (ASMC) systems based on the beam and bristle models. The PI control system cannot maintain the reference slip ratio during the braking time in the presence of system uncertainties. However, the ASMC system maintains the reference slip ratio during the braking time, despite system uncertainties. Using the ASMC based on the bristle model in place of the PI control based on the beam model yields 28% improvement in the wheel slip.

TABLE 5 The performance of the PI and ASMC systems based on the beam and bristle models. λ and λ_d denote the slip ratio and reference slip ratio, respectively. The ASMC based on the bristle model has the most effective performance and robustness for coping with system uncertainties.

	PI		ASMC	
Performance	Beam model	Bristle model	Beam model	Bristle model
Braking distance	700 m	682 m	607 m	581 m
Braking time	59.5 s	58.9 s	55.3 s	50.7 s
Expended braking energy	1.77×10^{-7} kN-m	1.77×10^{-7} kN-m	1.71×10^{-7} kN-m	1.71×10^{-7} kN-m
Mean value of the absolute error between λ and λ_d	0.0378	0.0351	0.0354	0.0272

REFERENCES

[1] S. Kadowaki, K. Ohishi, S. Yasukawa, and T. Sano, "Anti-skid re-adhesion control based on disturbance observer considering air brake for electric commuter train," in *Proc. 8th IEEE Int. Workshop Advanced Motion Control*, Kawasaki, Japan, 2004, pp. 607–612.

[2] S. Shirai, "Adhesion phenomena at high-speed range and performance of an improved slip-detector," *Q. Rep. Railw. Tech. Res. Inst.,* vol. 18, no. 4, pp. 189–190, 1977.

[3] I.P. Isaev and A.L. Golubenko, "Improving experimental research into adhesion of the locomotive wheel with the rail," *Rail Int.*, vol. 20, no. 8, pp. 3–10, 1989.

[4] T. Ohyama, "Some basic studies on the influence of surface contamination on adhesion force between wheel and rail at high speed," *Q. Rep. Railw. Tech. Res. Inst.*, vol. 30, no. 3, pp. 127–135, 1989.

[5] K. Ohishi, K. Nakano, I. Miyashita, and S. Yasukawa, "Anti-slip control of electric motor coach based on disturbance observer," in *Proc. 5th IEEE Int. Workshop Advanced Motion Control*, Coimbra, Portugal, 1998, pp. 580–585.

[6] T. Watanabe and M. Yamashita, "A novel anti-slip control without speed sensor for electric railway vehicles," in *Proc. 27th Annu. Conf. IEEE Industrial Electronics Society*, Denver, CO, 2001, vol. 2, pp. 1382–1387.

[7] M.C. Wu and M.C. Shih, "Simulated and experimental study of hydraulic anti-lock braking system using sliding-mode PWM control," *Mechatronics*, vol. 13, 2003, pp. 331–351.

[8] A. Kawamura, T. Furuya, K. Takeuchi, Y. Takaoka, K. Yoshimoto, and M. Cao, "Maximum adhesion control for Shinkansen using the tractive force tester," in *Proc. IEEE Power Conversion Conf.*, Osaka, Japan, 2002, vol. 1, pp. 567–572.

[9] Y. Ishikawa and A. Kawamura, "Maximum adhesive force control in super high speed train," in *Proc. IEEE Power Conversion Conf.*, Nagaoka, Japan, 1997, vol. 2, pp. 951–954.

[10] Y. Takaoka and A. Kawamura, "Disturbance observer based adhesion control for Shinkansen," in *Proc. 6th IEEE Int. Workshop Advanced Motion Control*, Nagoya, Japan, 2000, pp. 169–174.

[11] C. Canudas de Wit, K.J. Astrom, and P. Lischinsky, "A new model for control of systems with friction," *IEEE Trans. Automat. Contr.*, vol. 40, no. 3, pp. 419–425, 1995.

[12] T.A. Johansen, I. Petersen, J. Kalkkuhl, and J. Lüdemann, "Gain-scheduled wheel slip control in automotive brake systems," *IEEE Trans. Contr. Syst. Technol.*, vol. 11, no. 6, pp. 799–811, 2003.

[13] L. Li, F.Y. Wang, and Q. Zhou, "Integrated longitudinal and lateral tire/road friction modeling and monitoring for vehicle motion control," *IEEE Trans. Intell. Transport. Syst.*, vol. 7, no. 1, pp. 1–19, 2006.

[14] M. Basset, C. Zimmer, and G.L. Gissinger, "Fuzzy approach to the real time longitudinal velocity estimation of a FWD car in critical situations," *Vehicle Syst., Dyn.*, vol. 27, no. 5 & 6, pp. 477–489, 1997.

[15] L. Alvarez, J. Yi, R. Horowitz, and L. Olmos, "Dynamic friction model-based tire-road friction estimation and emergency braking control," *ASME J. Dyn. Syst., Meas. Contr.*, vol. 127, no. 1, pp. 22–32, 2005.

[16] J. Yi, L. Alvarez, and R. Horowitz, "Adaptive emergency braking control with underestimation of friction coefficient," *IEEE Trans. Contr. Syst. Technol.*, vol. 10, no. 3, pp. 381–392, 2002.

[17] F. Gustafsson, "Slip-based tire-road friction estimation," *Automatica*, vol. 33, no. 6, pp. 1087–1099, 1997.

[18] S. Kumar, M.F. Alzoubi, and N.A. Allsayyed, "Wheel/rail adhesion wear investigation using a quarter scale laboratory testing facility," in *Proc. 1996 ASME/IEEE Joint Railroad Conf.*, Oakbrook, IL, 1996, pp. 247–254.

[19] X.S. Jin, W.H. Zhang, J. Zeng, Z.R. Zhou, Q.Y. Liu, and Z.F. Wen, "Adhesion experiment on a wheel/rail system and its numerical analysis," *IMechE J. Eng. Tribology*, vol. 218, pp. 293–303, 2004.

[20] W. Zhang, J. Chen, X. Wu, and X. Jin, "Wheel/rail adhesion and analysis by using full scale roller rig," *Wear*, vol. 253, no. 1, pp. 82–88, 2002.

[21] H. Sakai, *Tire Engineering* (in Japanese). Tokyo, Japan: Grand Prix Publishing, 1987.

[22] J.J. Kalker and J. Piotrowski, "Some new results in rolling contact," *Veh. Syst. Dyn.*, vol. 18, no. 4, pp. 223–242, 1989.

[23] J.J. Kalker, *Three-Dimensional Elastic Bodies in Rolling Contact*. Norwell, MA: Kluwer Academic, 1990.

[24] K. Knothe and S. Liebelt, "Determination of temperatures for sliding contact with applications for wheel-rail systems," *Wear*, vol. 189, pp. 91–99, 1995.

[25] P. Dahl, "Solid friction damping of mechanical vibrations," *AIAA J.*, vol. 14, no. 12, pp. 1675–1682, 1976.

[26] C. Canudas de Wit and P. Tsiotras, "Dynamic tire friction models for vehicle traction control," in *Proc. 38th IEEE Conf. Decision Control*, Phoenix, AZ, 1999, pp. 3746–3751.

[27] C. Canudas de Wit and P. Lischinsky, "Adaptive friction compensation with partially known dynamic model," *Int. J. Adapt. Contr. Signal Process.*, vol. 11, pp. 65–80, 1997.

[28] G. Charles and R. Goodall, "Low adhesion estimation," in *Proc. Institution Engineering Technology Int. Conf. Railway Condition Monitoring*, Birmingham, UK, 2006, pp. 96–101.

[29] Y. Shimizu, K. Ohishi, T. Sano, S. Yasukawa, and T. Koseki, "Anti-slip/skid re-adhesion control based on disturbance observer considering bogie vibration," in *Proc. 4th Power Conversion Conf.*, Nagoya, Japan, 2007, pp. 1376–1381.

[30] N. Patel, C. Edwards, and S.K. Spurgeon, "A sliding mode observer for tyre friction estimation during braking," in *Proc. 2006 American Control Conf.*, Minneapolis, Minnesota, 2006, pp. 5867–5872.

[31] S. Uchida, "Brake of railway III," *Japan Train Oper. Assoc. J. Brake of Railway Vehicles*, vol. 3, pp. 25–28, 2001.

[32] W. Gao and J.C. Hung, "Variable structure control of nonlinear systems," *IEEE Trans. Ind. Electron.*, vol. 40, no. 1, pp. 45–55, 1993.

[33] K.J. Lee, H.M. Kim, and J.S. Kim, "Design of a chattering-free sliding mode controller with a friction compensator for motion control of a ball-screw system," *IMechE J. Syst. Contr. Eng.*, vol. 218, pp. 369–380, 2004.

[34] H.K. Khalil, *Nonlinear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

AUTHOR INFORMATION

Sung Hwan Park received the B.S. and M.S. in precision mechanical engineering from Pusan National University, Korea, in 1990 and 1992, respectively, the Ph.D. in mechanical engineering from Pusan National University in 1996, and the Dr. Eng. in control systems engineering from Tokyo Institute of Technology in 2005. He is a research assistant professor in the School of Mechanical Engineering, Pusan National University. His research interests include hydraulic control systems, design of construction machinery, and brake systems for railway rolling stocks.

Jong Shik Kim (jskim@pusan.ac.kr) received the B.S. in mechanical design and production engineering from Seoul National University in 1977. He received the M.S. in mechanical engineering from Korea Advanced Institute of Science and Technology in 1979 and the Ph.D. in mechanical engineering from MIT in 1987. He is a professor in the School of Mechanical Engineering, Pusan National University, Korea. His research interests include dynamics and control of vehicle systems. He can be contacted at the School of Mechanical Engineering, Pusan National University, Busan 609-735, South Korea.

Jeong Ju Choi received the B.S. in mechanical engineering from Dong-A University in 1997. He received the M.S. and Ph.D. in mechanical and intelligent systems engineering from Pusan National University in 2001 and 2006, respectively. He is a postdoctoral fellow in the Department of Industrial and Manufacturing Systems Engineering at the University of Michigan-Dearborn. His research interests include dynamics and control of field robot systems.

Hiro-o Yamazaki received the B.S. and Ph.D. in industrial mechanical engineering from Tokyo University of Agriculture and Technology in 1992 and 2005, respectively. He is an associate professor in the Department of Mechanical Engineering, Faculty of Engineering, Musashi Institute of Technology in Japan. His research interests include dynamics and control of brake systems for railway rolling stocks.

Friction Modeling in Linear Chemical-Mechanical Planarization

JINGANG YI

PROCESS MONITORING FOR WAFER POLISHING IN SEMICONDUCTOR MANUFACTURING

Chemical-mechanical planarization (CMP) is an abrasive process for polishing the surfaces of wafer flats in integrated-circuit (IC) fabrication. ICs are made on wafers and stacked up in three-dimensional layer structures; for details, see "Overview of Semiconductor Manufacturing" in [1]. The feature size or critical dimension of IC devices is determined by the pattern lithography process. A planarized topography of the device surface is critical for the lithography process to achieve a precise patterned feature at the submicron level. CMP is the enabling technology for achieving global planarization in semiconductor manufacturing. For example, interlevel dielectric (ILD) and intermetal dielectric (IMD) are widely used microstructures in memory and logic integrated devices. Figure 1 illustrates the ILD/IMD cross-section structures with and without CMP. Nonplanarization of ILD/IMD structures accumulates from the bottom levels to the upper levels during multilayer IC fabrication. Such nonplanarization layers not only limit the depth of focus of the high-resolution lithography process but also deteriorate the performance of ICs [2].

During CMP, the device side of the wafer is pushed down by a wafer carrier against a moving polishing pad to achieve surface planarization; see Figure 2. Chemical slurry is poured onto the polishing pad to assist in material removal and surface planarization. The chemical reaction between the slurry fluid and the film as well as the interac-

tion between the particles in the slurry and the wafer surface aid material removal during the polishing process. The slurry fluids also help to remove the ground-off materials and by-products. To ensure consistent performance, the roughness of the polishing pad surface is maintained by a pad-conditioning system, shown in Figure 2(b). Three types of CMP polishers are used in the semiconductor manufacturing industry, namely, rotational, orbital, and linear polishing machines. Figure 2(a) shows a schematic of a rotational CMP configuration, while (b) shows a linear CMP configuration. For rotary CMP machines, both the wafer carrier and the polishing pad rotate, while for linear CMP machines the polishing pad moves linearly across the wafer surface and is supported by a pair of rollers.

CMP involves both mechanical and chemical interaction between solids and fluids, while CMP performance is determined by several mechanical and chemical process parameters. These parameters include process and machine configurations such as polishing down force, consumable material-related parameters, such as slurry pH values and temperature, and wafer-related parameters, such as wafer film materials and topography [2]. Various metrics are defined to quantify CMP performance; see "How to Quantify CMP Performance." From the system's point of view, the process input configurations and output performance are illustrated in Figure 3. The input parameters include slurry, pad, wafer, and equipment and process configurations. The output performance parameters are defined at various scales, such as wafer, die, feature, and particle scale; see "CMP for Shallow Trench Isolation Patterned Wafers."

Two modeling approaches, statistical and physical models, are used to analyze CMP. The statistical modeling approach uses statistics methods to construct a data-based mathematical model between the process inputs, such as process configurations, and the process outputs, such as the material removal rate. Such a modeling approach has been widely used for advanced run-to-run process control of CMP [3]–[6]. The physical modeling approach, on the other hand, investigates interactions among the wafer, polishing pad, and slurry fluids, at both the micro and macroscales. One of the advantages of the physical modeling approach over the statistical modeling approach is that the former gives physical interpretations and can be used as a mathematical foundation for real-time process monitoring and control.

Although polishing is a major manufacturing process, no single analytical model provides a precise relationship between CMP input parameters and output performance parameters. Poor process understanding and a lack of effective in situ process sensors due to dynamic and harsh environmental process conditions are the main challenges to developing analytical models. Modeling of the polishing process can be traced back to the beginning of the last century [7]. In recent years, various mathematical models have been developed to quantitatively understand the material removal mechanism in CMP [2]. In these models, three types of wafer/pad contact regimes are assumed, namely, direct solid-to-solid contact, semidirect contact, and lubrication and hydrodynamic contact. The thickness of the thin film between the wafer and pad determines the regime to which a particular process setup belongs. Figure 4 illustrates the relationship among contact regimes, film thickness, and polishing parameters, such as slurry viscosity, relative velocity between the wafer and pad, and applied wafer/pad pressure.

Friction between the wafer and the polishing pad contributes significantly to CMP performance, such as the
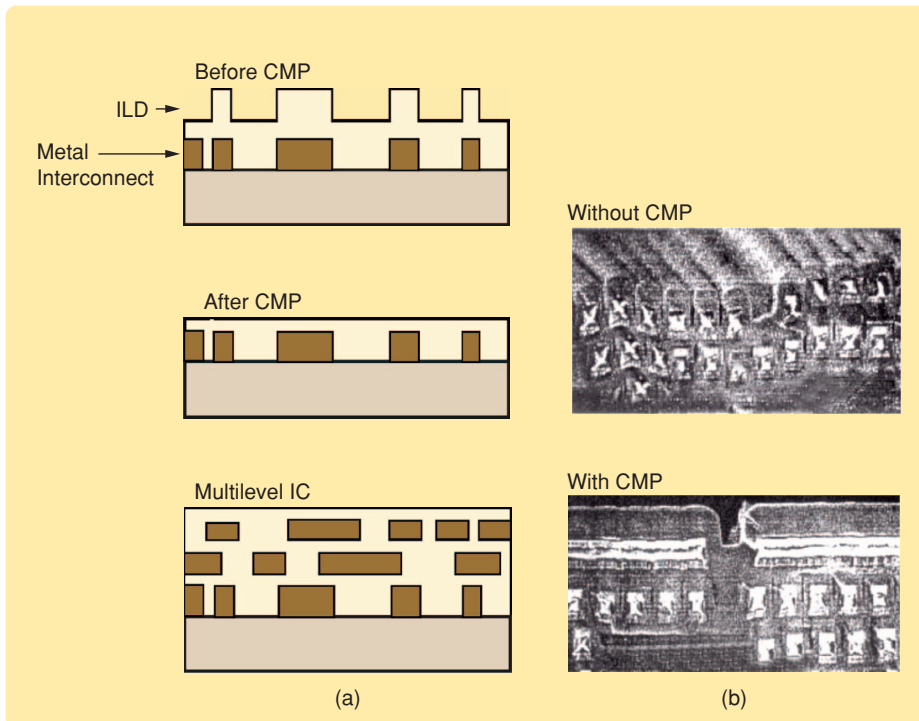
FIGURE 1 A schematic diagram showing multilevel integrated circuit structures. Interlevel dielectric (ILD) and intermetal dielectric (IMD) are widely used structures for constructing multilevel integrated circuits. Metal layers/lines are marked as red color rectangular blocks. (a) The process flow of chemical-mechanical planarization (CMP) for ILD/IMD structures and (b) a cross-section comparison of an ILD/IMD structure with and without CMP. (From [2], used with kind permission of Springer-Science+Business Media.)

material removal rate [8]–[11]. In [12], the coefficient of friction for a prototype of a rotary CMP polisher is measured under various polishing parameters, such as the polishing table's rotational speed and polishing down force. The study in [9] concludes that the material removal rate for copper CMP is partially driven by the coefficient of friction. In [10] and [11], it is demonstrated experimentally that the material removal rate is proportional to the wafer/pad friction forces for both oxide- and copper-wafer CMP. In [13], a mechanism and method is designed to measure the friction forces in real time. The relationship between the wafer/pad friction forces and process configurations is explored, both analytically and experimentally, in [14] and [15]. The focus in [14] and [15] is on the mechanical friction effects at wafer, die, and feature scale for oxide CMP. These analytical models provide a framework for studying wafer/pad friction and thus enable the in situ monitoring and control of CMP.

In this article, we develop an analytical model of the relationship between the wafer/pad friction and process configuration. We also provide experimental validation of this model for in situ process monitoring. CMP thus demonstrates that the knowledge and methodologies developed for friction modeling and control can be used to advance the understanding, monitoring, and control of semiconductor manufacturing processes. Meanwhile, relevant issues and challenges in real-time monitoring of CMP are presented as sources of future development.

This article is organized as follows. In the following section, we describe the linear CMP, which we focus on in this article. Then we discuss wafer/pad friction modeling in three steps. In the first step, we present a wafer/pad friction model without considering pad-conditioning and patterned-wafer topography effects. The pad-conditioning effect on wafer/pad friction is investigated in the next step. Then we discuss the effect of patterned-wafer topography on wafer/pad friction characteristics. Finally, experimental model validations are presented, followed by some application examples.

LINEAR CHEMICAL-MECHANICAL PLANARIZATION

The linear-planarization mechanism is different from that of a rotary polisher. In the linear CMP setup, the polishing pad moves linearly against the rotating wafer. A circular air-bearing system, shown in Figure 5, is used to adjust the polishing uniformity. Compressed air between the polishing pad and the air platen supports the moving polishing pad. The air zones on the platen consist of small cocentered holes located at various radii as shown in
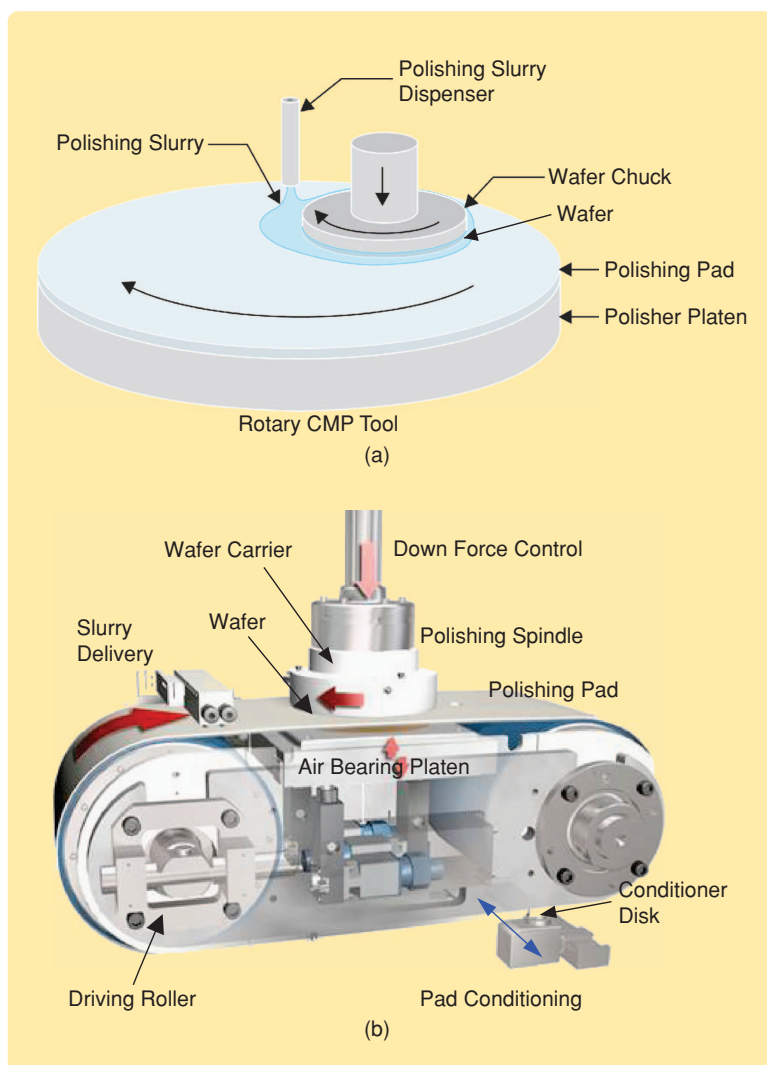

FIGURE 2 Typical configurations of the chemical-mechanical planarization (CMP) modules. In (a) the rotational configuration, both the wafer carrier and polishing pad are rotating. In (b) the linear configuration, the wafer carrier rotates, while the polishing pad moves linearly across the wafer surface. (Picture used courtesy of Lam Research Corporation.)

Figure 5(b). By tuning the air pressure in each air zone on the platen and by adjusting the air gap between the platen and the polishing pad, we can change the polishing pad deformation and thus adjust the wafer polishing uniformity [6]. Moreover, the air-bearing tuning mechanism decouples control of the polishing process uniformity from control of the material removal rate [16]. Compared with rotary CMP tools, linear planarization provides a wide range of polishing pad speeds and polishing pressures and therefore increases the process throughput as well as the planarization performance [16].

In CMP, the surface of the polish pad must be maintained at a specified roughness level in order to maintain consistent performance. Conditioning the pad is an effective method for maintaining the pad surface roughness [17]. In practice, a moving conditioner disk is pushed against the moving pad. Hundreds of small diamonds, shown in Figure 6(a), are usually mounted on top of the conditioner disk surface. As the disk moves across the pad, the disk scratches the porous pad surface, which sustains the pad surface roughness level. Figure 6(b) illustrates a linear conditioner system. The linear conditioner disk moves in a straight line across the polishing pad in a direction that is perpendicular to the pad movement direction.

FRICTION MODELING WITHOUT PAD CONDITIONING

The friction force between the wafer and polishing pad during polishing depends highly on the relative velocity distribution and wafer/pad surface characteristics. In this section, we discuss nonpatterned wafer/pad friction modeling without considering pad-conditioning effects.

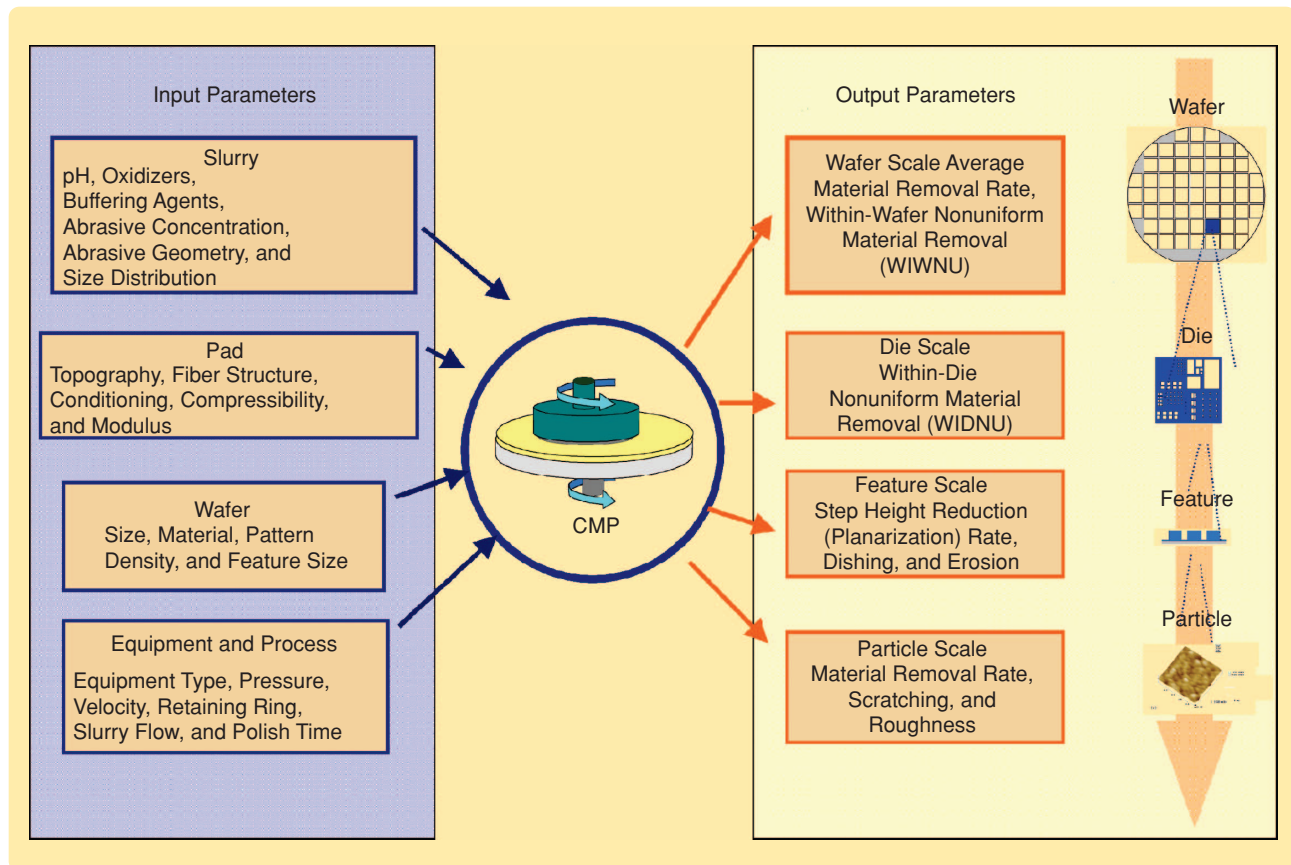**FIGURE 3** Input and output chemical-mechanical planarization (CMP) process parameters. CMP is a complex process that involves both mechanical and chemical interactions between solids and fluids. Multiple input factors, such as properties and configurations of the slurry, polishing pad, wafer, and polishing equipment, affect CMP output performance. (From [2], used with kind permission of Springer-Science+Business Media.)

The polishing pad is made of layers of polyurethane. Figure 7 shows a scanning electron microscope picture of the pad cross section and a schematic of the wafer/pad contact. During polishing, the wafer is supported by the pad pores [18], [19]. Due to pad conditioning, numerous broken pores are created on the pad surface, allowing slurry to accumulate inside the pores. When a wafer is pushed against the moving pad, slurry particles are squeezed into the pad pores, which helps remove the thin films on the wafer surface.

Figure 8 illustrates the kinematics of the rotating wafer on the moving polishing pad. The polishing pad speed and wafer rotational speed are denoted by v_B and ω, respectively. The position of an arbitrary point A on the wafer surface is denoted by $\mathbf{r} = (r\cos\phi)\mathbf{i} + (r\sin\phi)\mathbf{j}$ in the x-y coordinate system, where r is the distance from the point A to the origin O, ϕ is the angle between the vector $\overrightarrow{OA}$ and the x-axis, and $\mathbf{i}$ and $\mathbf{j}$ are the unit vectors along the x- and y-axis directions, respectively. The relative velocity $\mathbf{v}_{\mathrm{rel}}$ of the point A

How to Quantify CMP Performance

Several specifications are widely used in the semiconductor industry to quantify CMP performance. Here we introduce a few commonly used metrics [2].

Material removal rate (MRR) is the amount of film material removal (in thickness) per unit time. In practice, wafer film thicknesses are measured at a set of specified locations before and after polishing. We denote the film thicknesses before and after polishing at the ith location as l_i^{pre} and l_i^{pst}, respectively. Denoting the polishing time ΔT, the material removal rate MRR_i at the ith location is

$$\mathrm{MRR}_i = \frac{l_i^{\mathrm{pre}} - l_i^{\mathrm{pst}}}{\Delta T}.$$

The CMP removal rate is then given by the average material removal rate $\mathrm{MRR}_{\mathrm{avg}}$ at all N measurement locations

$$\mathrm{MRR}_{\mathrm{avg}} = \frac{\sum_{i=1}^{N} \mathrm{MRR}_i}{N}.$$

Within-wafer nonuniformity (WIWNU) describes how CMP performs uniformly across the wafer. We denote the measured material removal rate at the ith location as MRR_i, $i = 1, \ldots, N$. WIWNU is defined as

$$\mathrm{WIWNU} = \frac{\mathrm{MRR}_{\max} - \mathrm{MRR}_{\min}}{2\mathrm{MRR}_{\mathrm{avg}}} \times 100\%,$$

where $\mathrm{MRR}_{\max} = \max_{1 \le i \le N} \mathrm{MRR}_i$ and $\mathrm{MRR}_{\min} = \min_{1 \le i \le N} \mathrm{MRR}_i$. WIWNU is also approximated as

$$\mathrm{WIWNU} = \frac{6\sigma_{\mathrm{MRR}}}{2\mathrm{MRR}_{\mathrm{avg}}} = \frac{3\sigma_{\mathrm{MRR}}}{\mathrm{MRR}_{\mathrm{avg}}} \times 100\%,$$

where σ_{MRR} is the standard deviation of the MMR measurements at all N locations across the wafer.

Dishing and *erosion* typically refer to copper dishing and oxide erosion, respectively, in a copper damascene process. Copper is increasingly being used to replace aluminum (Al) as the interconnect metal in ICs due to copper's lower resistivity. However, it is difficult to use a dry or wet etch process to pattern copper films. Instead, a copper damascene process is used. In the copper damascene process, copper is deposited on patterned dielectric-film trenches to make interconnect lines, and CMP is needed to remove the copper film on the dielectric films. Figure S1 illustrates an ideal and actual topography after the copper damascene CMP [2]. In Figure S1, copper lines, both wide and thin, are marked in red. Ideally, when CMP stops, the copper film is at the same level

as the oxide film. However, due to the fact that copper material is removed faster than oxide material, the oxide film thickness is reduced faster in regions of dense, thin copper lines. The amount of oxide thickness reduction at the dense copper line region, shown in Figure S1, is defined as erosion. For wider copper lines, the copper is removed faster than the oxide film. The resulting copper thickness reduction relative to its neighboring oxide films, shown in Figure S1, is called dishing.

The oxide-erosion and copper-dishing phenomenon is undesirable because oxide layer thinning leads to larger parasite capacitance between interconnection metal layers, while copper-layer thinning increases the resistance of the copper lines. Therefore, smaller erosion and dishing are more desirable for copper CMP. Erosion and dishing measurements, which are provided by a specialized metrology machine, are time consuming.

Other process output parameters include within-die nonuniformity (WIDNU), planarization rate, scratch, and surface roughness. More details for these parameters can be found in [2].
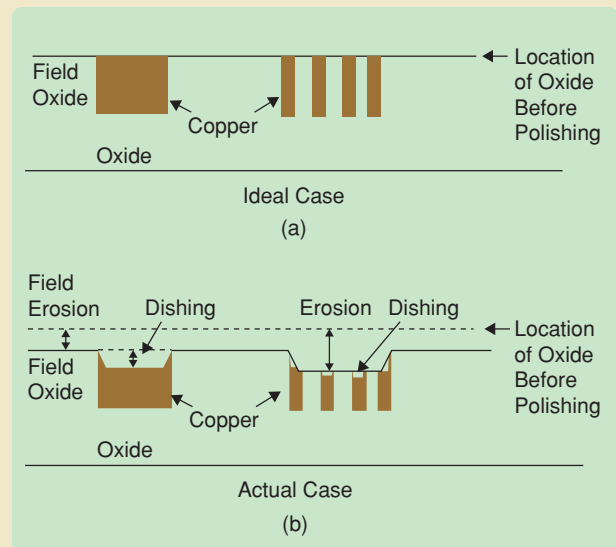
FIGURE S1 A schematic diagram of ideal and actual topography after copper damascene CMP. In (a) the ideal case, the oxide and copper layers are located at the same horizontal surface. However, the actual material removal rates for the copper and oxide films are different. (b) Copper dishing and oxide erosion are defined as the recess depths of the thick copper lines and the oxide film at thin copper lines, respectively. (From [2], used with kind permission of Springer-Science+Business Media.)

on the wafer with respect to the contact point on the pad is
$\mathbf{v}_{rel} = (-\omega r \sin\phi - v_B)\mathbf{i} + (\omega r \cos\phi)\mathbf{j}$.

For the wafer/pad friction force, we consider the LuGre friction model [20], which captures the friction interface between the contact surfaces. The friction force is considered as an aggregated effect of elastic contact bristles. A first-order nonlinear differential equation is used to describe the dynamics of the average deformation of the

CMP for Shallow Trench Isolation Patterned Wafers

IC devices are fabricated and duplicated by each individual die structure on the wafer. Figure S2 shows the shallow trench isolation (STI) devices at wafer, die, and feature scale. The wafer shown in Figure S2(a) is a 200-mm SK3 patterned wafer based on the MIT 961 mask layout. Figure S2(b) shows the die mask layout of the STI devices that are designed for CMP testing. Within a die of size 20 mm × 20 mm, a set of 25 4-mm × 4-mm subdie squares are fabricated with various densities and pitch lengths (pattern density and pitch length are defined below). Polishing a testing wafer gives a good overall indication of the CMP performance. Figure S2(c) shows the cross-section scanning electron microscope (SEM) image of the STI feature after the CMP process.

STI is a front-end enabling technology in submicron semiconductor manufacturing. STI provides an isolation layer between active CMOS devices, such as transistors. Figure S3 shows the cross-section SEM pictures before and after the STI CMP. To fabricate STI devices, silicon oxide and silicon nitride are sequentially deposited on the silicon wafer. Isolation trenches are then plasma etched into the silicon and then overfilled with silicon oxide by chemical vapor deposition. Next, the oxide layer is polished back to a planar surface using CMP, while removing as little of the nitride as possible. CMP stops at the targeted endpoint, shown in Figure S3(a). The trench width and the active film width are denoted by L_T and L_A, respectively. A pitch

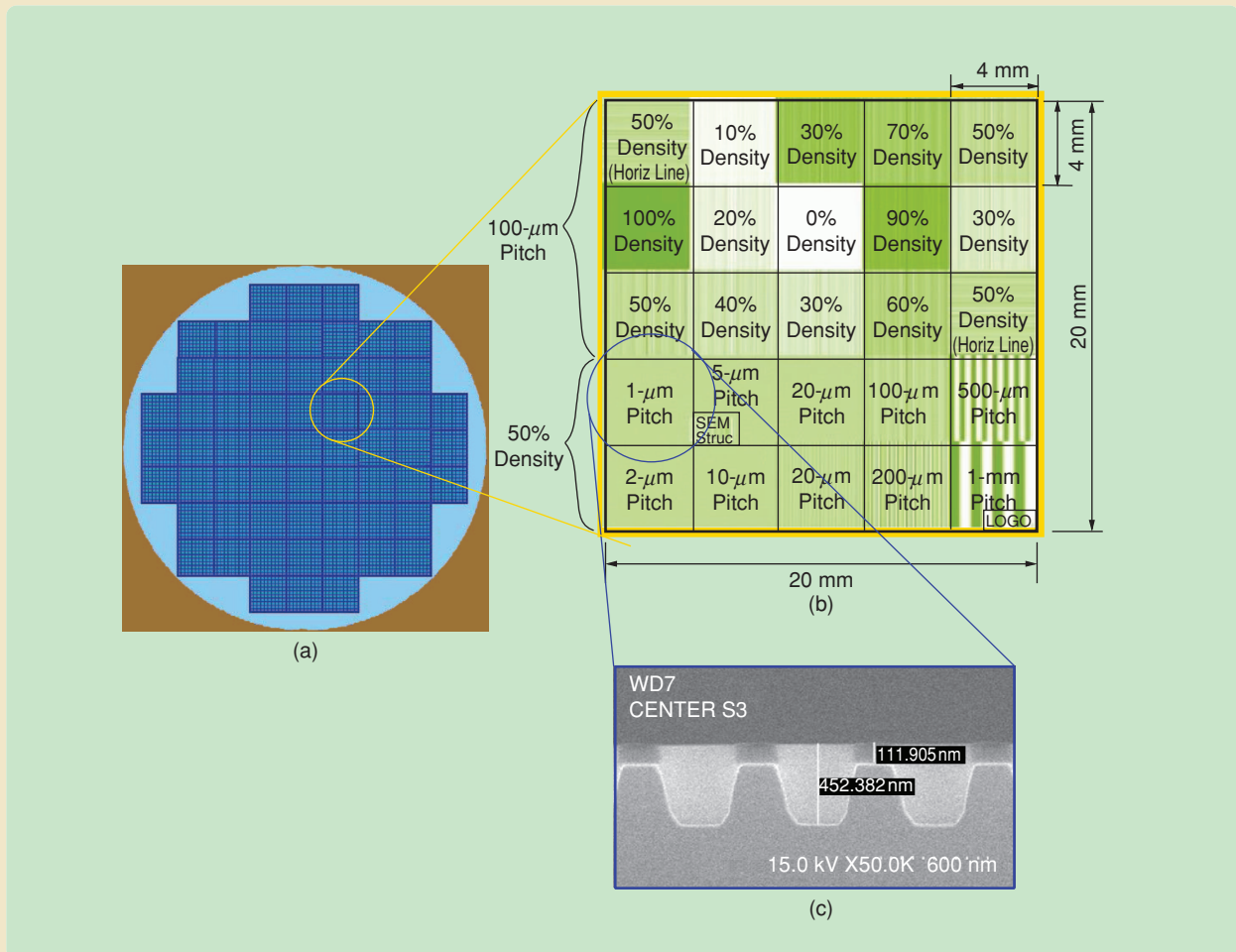FIGURE S2 Shallow-trench-isolation (STI) testing patterned wafer. (a) A 200-mm wafer top view and (b) a detailed mask layout (MIT 961) of each die. A total of 25 subdie structures are designed for a combination of various STI pattern densities and pitch lengths (b). (c) A cross-section scanning electron microscope picture of an STI feature with 50% pattern density and 500-μm pitch length after polishing and cleaning processes.

contact bristles. The friction force is then modeled as the resultant effect of bending and damping forces of the bristle deformation. To capture the effects of the normal force distribution as well as physical variations between the con-

tact surfaces, several distributed LuGre friction models are presented in [21] and [23] for automotive tire/road friction estimation and control. We consider the distributed LuGre friction model for wafer/pad interaction because of
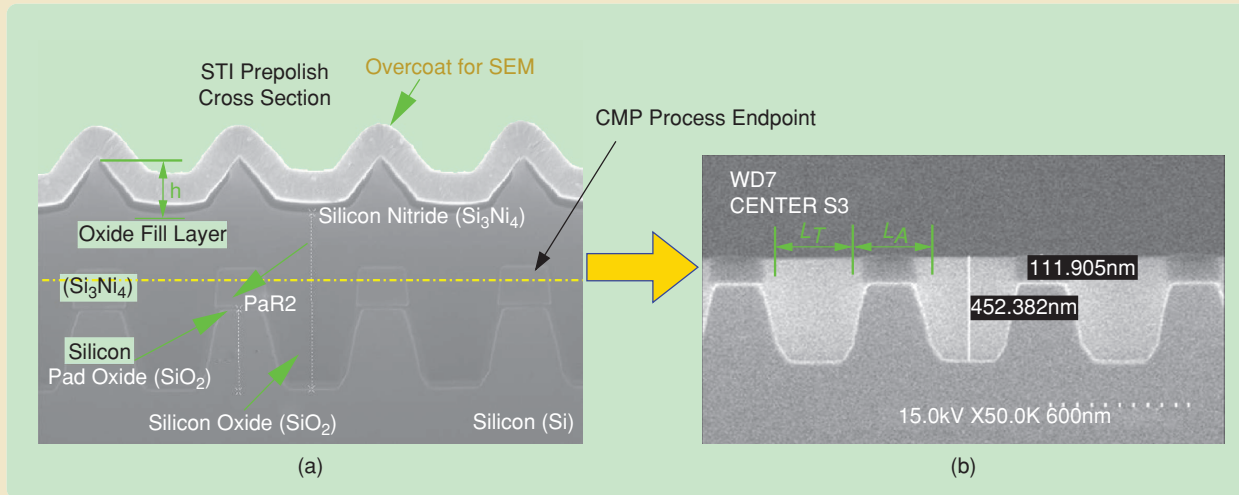

FIGURE S3 Comparison of scanning electron microscope images of a shallow-trench-isolation (STI) device before and after chemical-mechanical planarization (CMP). (a) A 50% pattern density STI device before CMP. CMP removes the deposited silicon oxide (SiO_2), and stops at the endpoint of the silicon nitride (Si_3Ni_4) layer, as indicated by the yellow dashed line. The STI feature after CMP and cleaning is shown in (b). The deposited silicon oxide film topography is planarized and ready for the next layer's processes.

consists of a trench area and an active area. The pitch length is $L_0 = L_A + L_T$. The STI pattern density ρ is defined as the ratio of the trench width over the pitch length, that is,

$$\rho = \frac{L_T}{L_0} = \frac{L_0 - L_A}{L_0} = 1 - \frac{L_A}{L_0}.$$

The step height h_S of the STI device is defined as the silicon oxide depth after the silicon oxide is deposited into the etched trench, shown in Figure S3(a). The formation of the trench is due to the silicon oxide conformation to the underlying device features.

A cluster tool, such as the CMP polisher, consists of many process modules; see Figure 16. The process provided by each module is configurable, and the process parameters on these modules are specified by users. The saved process setup and configuration are called process recipes. The CMP cluster tool computer system loads the recipes and executes the process sequences. Figure S4 illustrates a flow diagram of a typical process sequence on one polishing module. The polishing process starts with the ex situ pad-conditioning recipe and then the process recipe. Ex situ pad conditioning is executed before the wafer carrier touches down on the polishing pad. Ex situ pad conditioning is configured independently with the in situ pad conditioning and the polishing recipes. To keep the wafer/pad friction at the same roughness level, an ex situ pad-conditioning process is normally needed at the beginning of each process.
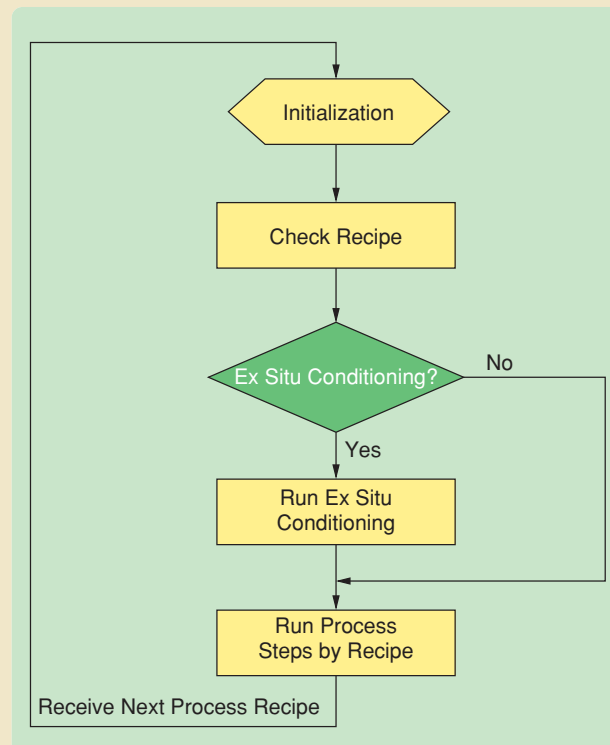

FIGURE S4 A schematic flow diagram of chemical-mechanical planarization (CMP) process sequences on one polishing module. Both ex situ and in situ pad conditioning are configurable. Pad conditioning, both ex situ and in situ, is used to tune CMP performance and maintain process consistency. (From [15].)

nonuniform relative velocity and normal force distribution across the wafer surface.

For nonpatterned wafers, the wafer/pad contact pressure p is uniformly distributed, where p is equal to the pressure p_0 applied by the polishing spindle system, that is, $p = p_0$. Let S_0 denote the wafer surface area, and consider a differential surface area dS_0 around the point A on the wafer. Using the distributed LuGre dynamic friction model, we obtain the friction force per unit area $\delta \mathbf{F}$ on dS_0 as

$$\frac{d\delta \mathbf{z}}{dt} = \mathbf{v}_{\text{rel}} - \theta \frac{\sigma_0 |\mathbf{v}_{rel}|}{g(\mathbf{v}_{\text{rel}})} \delta \mathbf{z}, \tag{1}$$

$$\delta \mathbf{F} = (\sigma_0 \delta \mathbf{z} + \sigma_1 \delta \dot{\mathbf{z}} + \sigma_2 \mathbf{v}_{\text{rel}}) \delta F_n, \tag{2}$$

where $\delta \mathbf{z}$ is the friction bristle deformation on dS_0, σ_0, σ_1, and σ_2 are the friction model parameters, and $\delta F_n = p_0$ is the normal pressure applied to dS_0. The model parameter $\theta > 0$ captures the physical interaction between the polishing pad and the wafer surface, for example, the variations
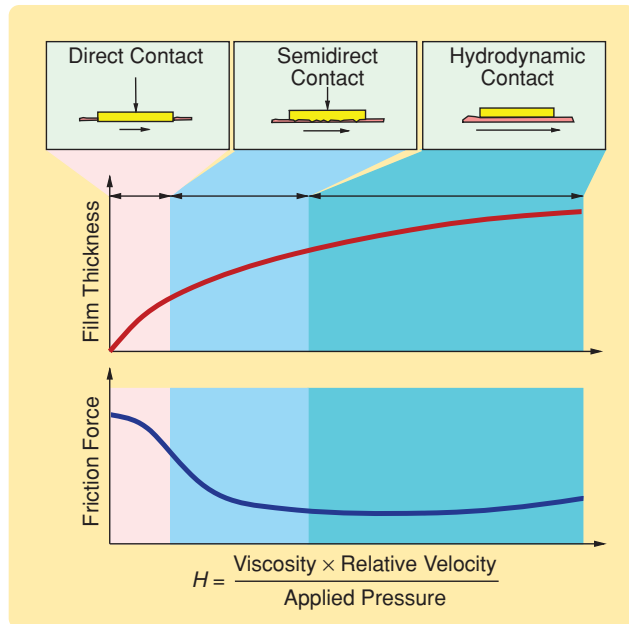

FIGURE 4 A schematic of slurry-film thickness, friction force, and various regimes for chemical-mechanical planarization. The wafer/pad friction and slurry-film thickness depend on three wafer/pad contact regimes, namely, direct, semidirect, and hydrodynamic contacts. This dependence is a function of the nondimensional Hersey number H. (From [2], used with kind permission of Springer-Science+Business Media.)

of the slurry fluid film thickness. The function $g(\mathbf{v}_{\text{rel}})$ is given by

$$g(\mathbf{v}_{\text{rel}}) = \mu_c + (\mu_s - \mu_c)e^{-\frac{|\mathbf{v}_{\text{rel}}|}{v_s}}, \tag{3}$$

where μ_c and μ_s are the Coulomb and static friction coefficients between the wafer and pad surface, respectively, and v_s is the Stribeck velocity.

The spindle and roller driving systems of the polishing module are shown in Figure 9. Since the wafer/pad relative velocity across the wafer is not symmetric with respect to the spindle rotational center, the spindle driving system must overcome the resultant friction torque generated by the wafer/pad interaction. Moreover, along the polishing pad movement direction, the wafer/pad friction force
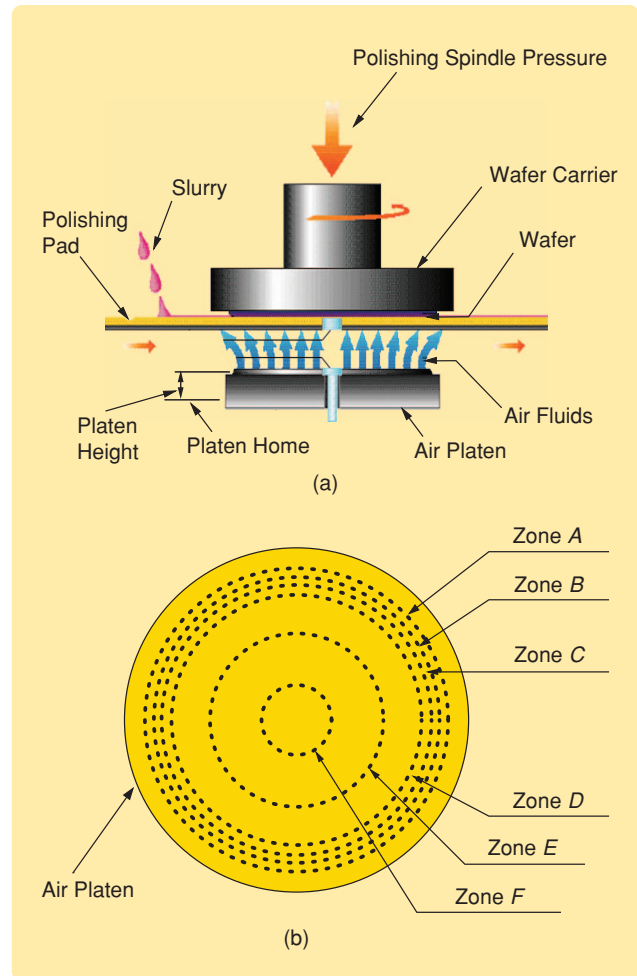
FIGURE 5 Linear planarization systems. (a) The air bearing system and (b) the air platen and its air zones. Controlled high-pressure air pushes through the platen air zones (A to F). By controlling the air zone pressures and adjusting the air gap between the platen and the polishing pad, the pressure distribution between the wafer and the polishing pad is tuned across the wafer surface. Linear planarization technology is used to decouple control of the polishing process uniformity from control of the material removal rate. (Pictures used courtesy of Lam Research Corporation.)

generates a friction torque around the roller rotational axis. The friction torques applied on the spindle and roller rotating axes are denoted by $\mathbf{M}^s$ and $\mathbf{M}^r$, respectively.

Spindle Friction Torque $\mathbf{M}^s$

The polishing spindle torque $\mathbf{M}^s$ with respect to the wafer center O is calculated as

$$\mathbf{M}^s = \int_{S_0} \mathbf{r} \times \delta \mathbf{F} dS_0. \tag{4}$$

We consider the steady-state friction torques by using the steady-state solutions of (1) and (2). Following the derivation in [14], we obtain

$$\delta \mathbf{F} = \left[\frac{g(\mathbf{v}_{\text{rel}})}{\theta |\mathbf{v}_{\text{rel}}|} + \sigma_2 \right] p_0 \mathbf{v}_{\text{rel}}. \tag{5}$$

Substituting (5) into (4) and using the approximation $g(\mathbf{v}_{\text{rel}}) \approx \mu_s - (\mu_s - \mu_c)(|\mathbf{v}_{\text{rel}}|/v_s)$ yields the torque magnitude M^s [14]. It is not possible, however, to obtain a closed-form relationship between the friction torque M^s and the polishing parameters. For most CMP, the pad speed v_B is fast, and the wafer carrier rotating speed ω is slow. By approximating $|\mathbf{v}_{\text{rel}}| \approx \omega r \sin \phi + v_B$, the friction torque M^s is obtained as [14]

$$M^s = \frac{1}{4} p_0 \pi R^4 \omega \left(\frac{\mu_s}{\theta v_B} + 2\sigma \right), \tag{6}$$

where $\sigma := \sigma_2 - (\mu_s - \mu_c/\theta v_s)$ and R is the wafer radius.

Roller Friction Torque $\mathbf{M}^r$

The friction torque applied to the driving roller rotating axis is calculated in the same fashion as for the spindle

(a)

(b)
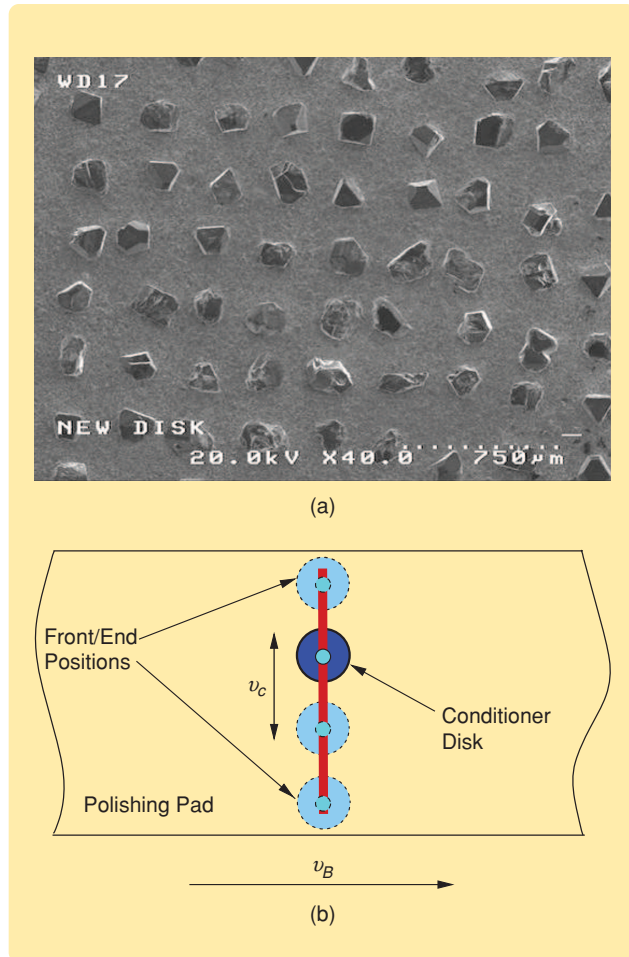
FIGURE 6 Linear pad-conditioning system. (a) A scanning electron microscope image of the conditioner disk top surface. A large number of small diamonds are mounted on the conditioner disk surface to facilitate pad conditioning. The conditioner disk rotates and moves along (b) a linear guide, which is orthogonal to the pad motion. The conditioner disk traveling speed v_c and down force against the polishing pad are individually controlled and configurable by process recipes. (Pictures used courtesy of Lam Research Corporation.)

(a)

(b)
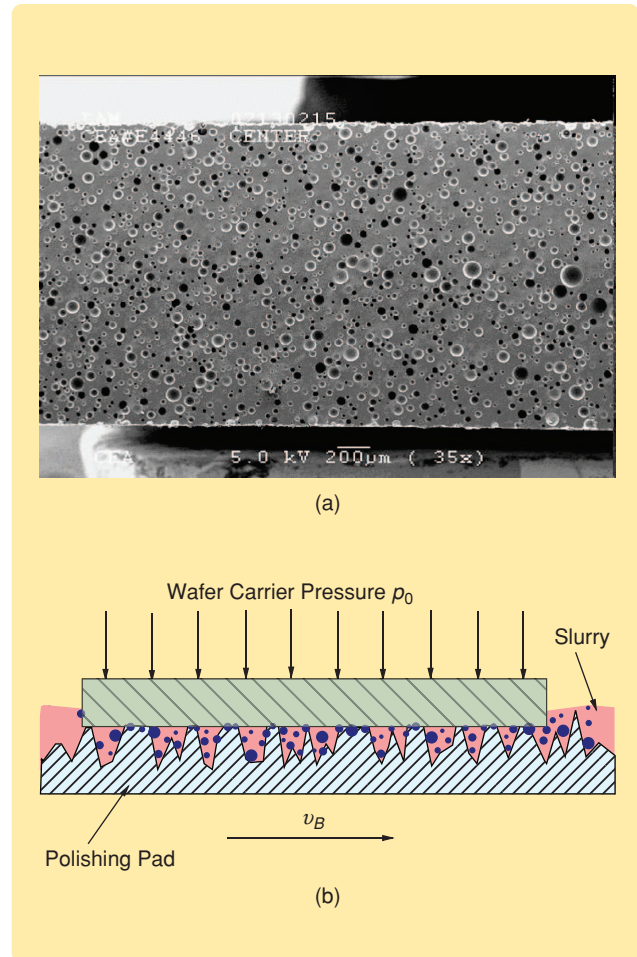
FIGURE 7 (a) A cross-section scanning electron microscope image of an IC 1000 pad. The polishing pad is made of a porous structure. Due to pad conditioning, numerous broken pores are created on the pad surface, allowing slurry to accumulate inside the pores. When a wafer is pushed against the moving pad, slurry particles are squeezed into the pad pores to help remove the thin film on the wafer surface (b). The porous material structure provides a stable contact surface under constant pad conditioning. (From [14].)
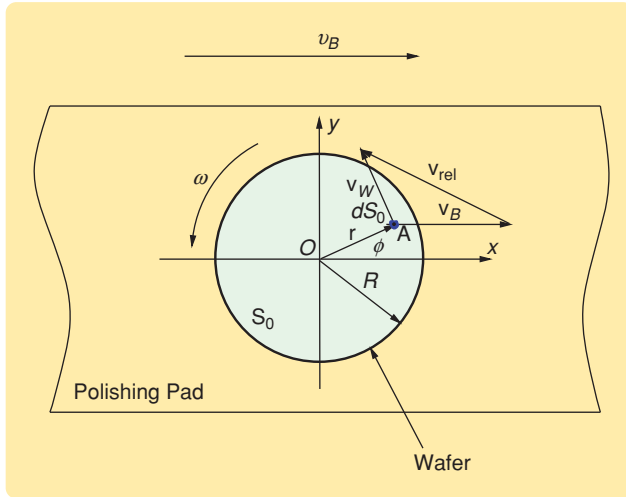
FIGURE 8 A schematic of the linear chemical-mechanical planarization wafer/pad kinematic relationship. The wafer rotates about its center O, while the polishing pad moves linearly with speed $\mathbf{v}_B$ across the wafer surface. The relative velocity $\mathbf{v}_{rel}$ of the differential element dS_0 on the wafer surface is the vector sum of the pad linear motion $\mathbf{v}_B$ and the wafer rotational motion $\mathbf{v}_W$, namely, $\mathbf{v}_{rel} = \mathbf{v}_B + \mathbf{v}_W$.

FIGURE 9 A schematic of the spindle and roller setup and force/torque calculation. The wafer rotates about the spindle axis with angular speed ω, while the polishing pad moves linearly across the wafer surface at speed $\mathbf{v}_B$. The friction torque $\mathbf{M}^s$ around the spindle rotating center as well as the friction torque $\mathbf{M}^r$ applied to the roller rotating axis are calculated by integrating the distributed friction forces across the wafer su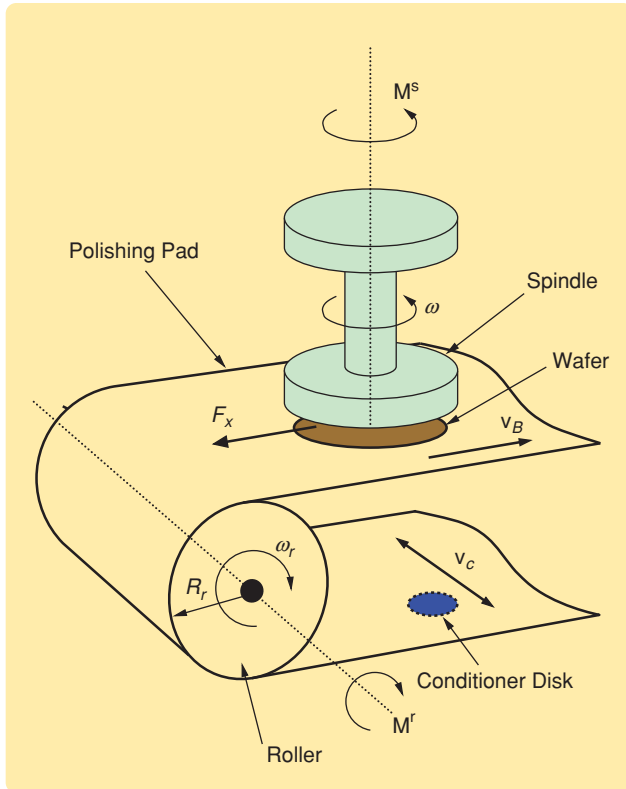rface. The magnitude of $\mathbf{M}^r$ is given by $M^r = F_x R_r$, where R_r is the roller radius and F_x is the resultant wafer/pad friction force along the polishing pad movement direction.

system. The wafer/pad friction force F_x along the direction of the pad motion is first obtained, and then the roller friction torque M^r due to the wafer/pad contact is determined [14]. Similar to the spindle-torque calculation, we obtain an approximation for M^r given by

$$M^r = p_0 \pi R^2 R_r \left(\frac{\mu_s}{\theta} + \sigma v_B \right), \qquad (7)$$

where R_r is the roller radius.

The model parameter σ satisfies $|\sigma| \ll 1$ because σ_2 is small and μ_s is slightly larger than μ_c, namely, $\mu_s - \mu_c \ll 1$. Therefore, the spindle torque M^s is proportional to the external pressure p_0, the wafer carrier speed ω, and the reciprocal of the pad speed v_B, while the roller torque M^r is proportional to the pressure p_0 but is not sensitive to variations of the wafer carrier speed ω and the pad speed v_B since σ is small. This theoretical conclusion is supported by experimental results in [15].

PAD CONDITIONING EFFECT ON FRICTION MODELS

Pad conditioning has a significant impact on CMP performance, especially the material removal rate. Pad conditioning restores the pad's effectiveness and reduces the decaying and glazing effects of the polishing pad. Figure 10 illustrates the microstructure comparison between an unused polishing pad surface and an unconditioned pad surface. Without pad conditioning, the pad surface cannot maintain its porous structure due to the dried slurry. Under pad conditioning, the wafer/pad friction coefficient is no longer constant and uniform across the entire pad surface. For the portion of the pad that is conditioned by the disk, namely, the shaded area shown in Figure 11, the friction coefficients are higher than the portion that is not conditioned. Therefore, we quantitatively capture the frictional characteristic difference between the portion of the pad that has been conditioned and the remaining portion that is not actively conditioned.

Figure 11 shows the kinematics of the conditioner system. The calculation of the trajectory of the conditioner disk center C on the pad is straightforward since the motion of the conditioner disk in the x- and y-axis directions is decoupled. We obtain the velocity of point C as $v_{Cx}(t) = v_B$ and $v_{Cy}(t) = \pm v_c$, where v_c is the conditioner-disk sweeping speed. The conditioner-disk position is restricted by $-L \leq y_C(t) \leq L$, where $2L$ is the maximum travel distance of the conditioner disk.

To quantify the impact of the conditioned pad portion on friction models, we consider the overlap of the trajectory of the conditioned portion of the polishing pad and the wafer surface. Figure 12 illustrates the kinematic relationship between a conditioned pad portion and the wafer surface. Let r_c denote the conditioner disk radius, and let $h(t)$ denote the distance from the

wafer center O to the center of the conditioner disk C. During one conditioning sweep, the conditioned pad portion, shown as the shaded area in Figure 12, moves across the wafer surface at a constant speed v_c. For simplicity, we assume that the end-limit positions of the conditioner disk are tangent with the wafer surface, and thus $h(t)$ changes from $R + r_c$ to 0 and then to $-(R + r_c)$. We first calculate the variations of the area of the conditioned pad portion overlapping with the wafer surface as a function of time t.

Let T_c denote the period of the conditioner-disk movement per sweep, and consider a two-sweep period when the conditioner disk moves from the rear to the front and then back to the rear positions. The conditioner-disk center position $h(t)$ is then

$$h(t) = \begin{cases} (R + r_c) - v_c t, & \text{if } 0 \le t < T_c, \\ -(R + r_c) + v_c(t - T_c), & \text{if } T_c \le t < 2T_c. \end{cases} \qquad (8)$$
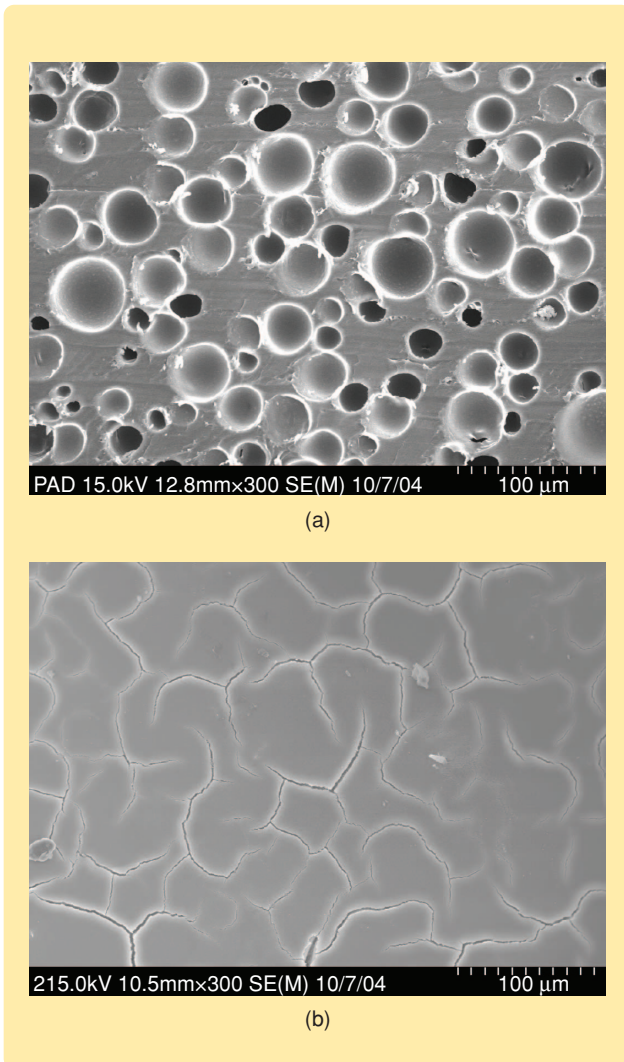

FIGURE 11 A schematic of the trajectory of the conditioned portion of the polishing pad. The portion of the polishing pad that has been conditioned, shown as the shaded area, has a higher friction coefficient than that of the remaining unconditioned portion due to recovery of the pad porous structure.

PAD 15.0kV 12.8mm×300 SE(M) 10/7/04   100 µm

(a)

215.0kV 10.5mm×300 SE(M) 10/7/04   100 µm

(b)

FIGURE 10 Comparison of scanning electron microscope images of the polishing pad surfaces. (a) An unused new pad surface and (b) a used pad surface without pad conditioning. Without pad conditioning, the pad surface is smoothed and glazed by slurries, in which case the pad becomes ineffective for material removal. (Reprinted from [27], with permission from Elsevier.)

FIGURE 12 The kinematic relationship between the conditioned pad portion and the wafer. The conditioned pad portion, shown as the shaded area, moves across the wafer surface. The conditioned pad portion has a higher friction coefficient than the portion of the polishing pad that the conditioner disk does not cover. When the polishing pad runs fast and the conditioner disk sweeps slowly across the polishing pad, the conditioned pad portion is approximately parallel to the pad movement direction. Therefore, the effect of the increased friction coefficient is estimated as the fraction of the shaded area over the entire wafer surface area. (From [14].)

Assuming that the polishing pad motion dominates the relative velocity between the wafer and pad, the increase in friction force due to pad conditioning is thus calculated as a fraction of the conditioned-pad-portion area, shown as the shaded area in Figure 12, out of the entire wafer surface area. For the spindle friction torque M^s, we consider both the magnitude and sign of the shaded area with respect to the wafer center O. Let $\Delta \mathbf{S}^s(t)$, $0 \le t < T_c$, denote the signed area of the conditioned pad portion within one rear-to-front conditioning sweep. The signed area $\Delta \mathbf{S}^s(t)$ depends on $h(t)$. For example, when $0 \le t < (2r_c/v_c)$, it follows that $R - r_c < h(t) \le R + r_c$. In this case, it follows from Figure 12 that $\alpha_2 = 0$ and $\alpha_1 = \cos^{-1}[(h(t) - r_c)/R]$. Therefore, the shaded area in Figure 12 is calculated as

$$\Delta \mathbf{S}^s(t) = \frac{2\alpha_1}{2\pi}\pi R^2 - \frac{1}{2}(2R\sin\alpha_1)(R\cos\alpha_1)$$

$$= R^2\left[\cos^{-1}\left(\frac{h - r_c}{R}\right) - \frac{h - r_c}{R}\sqrt{1 - \frac{(h - r_c)^2}{R^2}}\right].$$

$$(9)$$

Here, $h := h(t)$ for a simpler notation. Similarly, for $0 \le t \le T_c$, $\Delta \mathbf{S}^s(t)$ for the remaining cases is given by

$$\Delta \mathbf{S}^s(t) = \begin{cases} R^2\left[\cos^{-1}\left(\frac{h-r_c}{R}\right) - \frac{h-r_c}{R}\sqrt{1 - \frac{(h-r_c)^2}{R^2}}\right], \\ \qquad \text{if } 0 \le t < \frac{2r_c}{v_c}, \\[2mm] R^2\left\{\cos^{-1}\left(\frac{h-r_c}{R}\right) - \cos^{-1}\left(\frac{h+r_c}{R}\right) - \frac{1}{R}\left[(h - r_c)\right. \right. \\ \qquad \left.\left. \times \sqrt{1 - \frac{(h-r_c)^2}{R^2}} - (h + r_c)\sqrt{1 - \frac{(h+r_c)^2}{R^2}}\right]\right\}, \\ \qquad \text{if } \frac{2r_c}{v_c} \le t < \frac{R}{v_c}, \\[2mm] R^2\left\{\cos^{-1}\left(\frac{r_c-h}{R}\right) - \cos^{-1}\left(\frac{h+r_c}{R}\right) - \frac{1}{R}\left[(r_c - h)\right. \right. \\ \qquad \left.\left. \times \sqrt{1 - \frac{(h-r_c)^2}{R^2}} - (h + r_c)\sqrt{1 - \frac{(h+r_c)^2}{R^2}}\right]\right\}, \\ \qquad \text{if } \frac{R}{v_c} \le t < \frac{R+2r_c}{v_c}, \\[2mm] -R^2\left\{\cos^{-1}\left(\frac{|h|-r_c}{R}\right) - \cos^{-1}\left(\frac{|h|+r_c}{R}\right) \right. \\ \qquad - \frac{1}{R}\left[(|h| - r_c)\sqrt{1 - \frac{(|h|-r_c)^2}{R^2}} - (|h| + r_c)\right. \\ \qquad \left.\left. \times \sqrt{1 - \frac{(|h|+r_c)^2}{R^2}}\right]\right\}, \\ \qquad \text{if } \frac{R+2r_c}{v_c} \le t < \frac{2R-r_c}{v_c}, \\[2mm] -R^2\left[\cos^{-1}\left(\frac{|h|-r_c}{R}\right) - \frac{|h|-r_c}{R}\sqrt{1 - \frac{(|h|-r_c)^2}{R^2}}\right], \\ \qquad \text{if } \frac{2R-r_c}{v_c} \le t < T_c. \end{cases}$$

$$(10)$$

Due to symmetry, we obtain a similar formula for $T_c \le t < 2T_c$. In practice, v_c is calculated as $v_c = (2L/T_c) = (2(R + r_c)/T_c)$. The relationship $R \ge 3r_c$ holds for the polishing systems, and thus conditions in the fourth case in (10) are valid.

We are now ready to estimate the difference between the friction torque with and without pad conditioning. Suppose that after pad conditioning the wafer/pad friction coefficients, both Coulomb and static friction coefficients, are increased by the same amount $\gamma_\mu \ge 0$, namely,

$$\mu_{c\text{cond}} = (1 + \gamma_\mu)\mu_c, \quad \mu_{s\text{cond}} = (1 + \gamma_\mu)\mu_s, \qquad (11)$$

where $\mu_{c\text{cond}}$ and $\mu_{s\text{cond}}$ are the Coulomb and static wafer/pad friction coefficients, respectively, after pad conditioning. Since the increase of the friction coefficients happens only on the conditioned pad portion, the additional spindle torque due to pad conditioning is proportional to the fraction of the conditioned pad area. We thus obtain the spindle friction torque as

$$M^s_{\text{cond}}(t) = \left(1 + \gamma_\mu \frac{\Delta \mathbf{S}^s(t)}{\pi R^2}\frac{|h(t)|}{R}\right)M^s$$

$$= \left(1 + \gamma_\mu \gamma^s_{\text{cond}}(t)\right)M^s, \quad t \in [0, T_c), \quad (12)$$

where $\gamma^s_{\text{cond}}(t) := \Delta \mathbf{S}^s(t)|h(t)|/\pi R^3$ is the pad-conditioning factor at time t.

In (12) we consider the effect of only one conditioned pad trajectory within one pad rotational cycle. For CMP, the polishing pad runs at high speed, and the conditioner sweep speed is relatively slow. During one conditioner-disk sweep, the conditioned pad trajectory passes across the wafer surface several times. To capture the effect of multiple pad rotational cycles, we need to modify $\Delta \mathbf{S}^s(t)$ in (12) to $\Delta \mathbf{S}^s_e(t)$ as

$$\Delta \mathbf{S}^s_e(t) = \Delta \mathbf{S}^s(t) + \sum_{i=1}^{K}\beta_i\Delta \mathbf{S}^s(t - iT_B), \qquad (13)$$

where $0 < \beta_i < 1$, $i = 1, \dots, K$, are the forgetting factors, K is the number of pad-rotation cycles within one pad-conditioning sweep, $1 \le K = \lfloor v_B T_c/L_B \rfloor$ (where $\lfloor x \rfloor$ means the largest integer less than x), and L_B is the length of the polishing pad. The time constant $T_B = L_B/v_B$ is defined as the time period for the conditioned pad trajectory to travel one pad length. A total of T_c/T_B conditioned pad trajectories are generated on the polishing pad during one conditioning sweep. Intuitively, the calculation in (13) captures the impact of the multiple conditioned portions of the polishing pad on the friction models. The decaying wafer/pad friction coefficients, from $\mu_{c\text{cond}}$ to μ_c and from $\mu_{s\text{cond}}$ to μ_s, are represented by a set of forgetting factors β_i. Topologically, the conditioner-disk trajectory can be viewed as linear motion on a torus since

the polishing pad, when being cut apart along the conditioner disk movement direction, is of a rectangular shape, and the linear motion of the condition disk and the polishing pad are perpendicular to each other along the two sides of the rectangular pad. When the ratio $v_B T_c / L_B$ is an irrational number, the conditioner disk covers the entire pad surface after a large number of sweeps. The values of the forgetting factors β_i depend on the polishing and conditioner systems, and in practice these values are determined empirically.

We calculate the effect of pad conditioning on roller friction torque M^r in the same way. One difference is that the pad-conditioning effect always adds positive load on the roller motor, while, for the spindle torque, the extra load could be either positive or negative. Therefore, we obtain an unsigned area $\Delta \mathbf{S}^r(t)$ of the conditioned pad portion for the roller friction torque [14]. We modify (12) and (13) accordingly for the roller system torque $M^r_{cond}(t)$, while a similar factor γ^r_{cond} is introduced to model the pad-conditioning impact as

$$M^r_{cond}(t) = \left(1 + \gamma_\mu \gamma^r_{cond}(t)\right) M^r. \tag{14}$$

Figure 13 illustrates the spindle pad-conditioning factor $\gamma^s_{cond}(t)$ and roller pad-conditioning factor $\gamma^r_{cond}(t)$, respectively. In the calculation, we use $\beta_1 = 0.6$, $\beta_2 = 0.3$, $K = 2$, $T_c = 7$ s, $v_B = 200$ ft/min, and $L_B = 2.38$ m. From Figure 13, we notice different pad-conditioning effects on the spindle and roller torques. For both the spindle and roller systems, the extra friction torque fluctuations due to pad conditioning look like sinusoidal curves. However, for the spindle system, the oscillation period of the pad-conditioning factor $\gamma^s_{cond}(t)$ is $2T_c$ and the value of $\gamma^s_{cond}(t)$ is around zero, while for the roller system the period of the pad-conditioning factor $\gamma^r_{cond}(t)$ is T_c and the value of $\gamma^r_{cond}(t)$ is always nonnegative.

PATTERNED WAFER TOPOGRAPHY EFFECT
In this section, we consider the effect of the patterned-wafer film topography on the friction models. Since the local wafer/pad contact pressure decreases during patterned-wafer polishing, the friction forces between patterned wafer and the polishing pad also decrease. At the beginning of the polishing process, the local wafer/pad contact pressure $p = \delta F_n$ is higher than the external pressure p_0 because the wafer/pad contact area

$S(t)$ is much smaller than the wafer surface area S_0. Assuming that all dies on the wafer are identical, we obtain the average pressure p as

$$p = \delta F_n = \frac{p_0 S_0}{S(t)} = p_0 \gamma_{patt}(t), \tag{15}$$

where the pattern factor $\gamma_{patt}(t) := S_0 / S(t)$ captures the effect of evolution of the film topography of patterned wafers with time t. Since the pressure δF_n has a multiplicative effect in the LuGre friction model (2), using (12), (14), and (15), the spindle and roller torques are therefore modified as

$$M^s_{patt}(t) = \gamma_{patt}(t) \left[1 + \gamma_\mu \gamma^s_{cond}(t)\right] M^s,$$
$$M^r_{patt}(t) = \gamma_{patt}(t) \left[1 + \gamma_\mu \gamma^r_{cond}(t)\right] M^r, \tag{16}$$

respectively, for patterned-wafer polishing processes with pad conditioning.
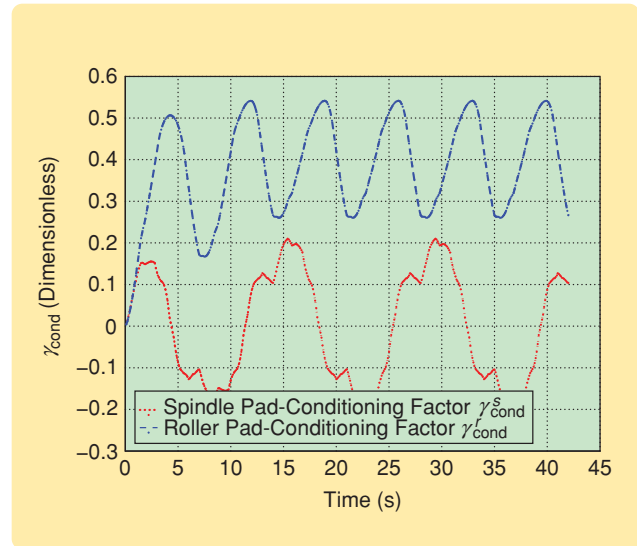
FIGURE 13 Friction torque pad-conditioning factor $\gamma_{cond}(t)$. Pad conditioning affects the friction torques of the polishing spindle and roller systems. However, the pad-conditioning factors for the spindle and roller systems have different characteristics. The period of the spindle pad-conditioning factor is twice the period of the roller pad-conditioning factor. Moreover, the magnitude of the spindle pad-conditioning factor is symmetric around zero, while the magnitude of the roller pad-conditioning factor is always positive. These differences are due to the distribution of the friction force across the wafer surface. (From [14].)

In the following, we use the shallow trench isolation (STI) patterned wafers to illustrate how we calculate the pattern factor $\gamma_{\text{patt}}(t)$; see "CMP for Shallow Trench Isolation Patterned Wafers" for a description of the STI polishing process.

Figure 14 provides a snapshot sequence of the cross-section film topography change during STI polishing. The initial trench step height is denoted by h_{S_0}, shown in Figure 14(a). Let h_c denote the contact height at which the polishing pad starts to touch the trench oxide surface, shown in Figure 14(b). Assuming $h_{S_0} \geq h_c$ when the process starts, the pad contacts only the active oxide area, and therefore the wafer/pad contact area $S(t)$ is equal to the total active oxide area across the wafer. When more oxide materials are removed, the step height h_S decreases, and the polishing pad starts to touch the trench oxide as shown in Figure 14(b). As the polishing process continues, the contact area $S(t)$ keeps increasing,

as shown in Figure 14(c). Finally, the active oxide area is completely removed, namely, $h_S = 0$, and the wafer/pad contact area is the entire wafer surface, namely, $S(t) = S_0$, shown in Figure 14(d).

We consider the step height topography evolution model in [24]. The pattern factor $\gamma_{\text{patt}}(t)$ is calculated as two cases. In Case I, $h_c \leq h_{S_0}$, and

$$
\gamma_{\text{patt}}(t) = \begin{cases} \frac{1}{1-\rho}, & 0 \leq t < t_c, \\ \frac{1}{1-\rho e^{-\frac{t-t_c}{\tau_B}}}, & t \geq t_c, \end{cases} \tag{17}
$$

where t_c is the time instant when the step height $h_S(t)$ reaches the contact height h_c, $\tau_B = (\rho h_c / \text{MRR}_B)$ is the polishing time constant for the film height h_c, ρ is the pattern density, and MRR_B is the oxide-wafer material removal rate. Next, in Case II, $h_c > h_{S_0}$, and
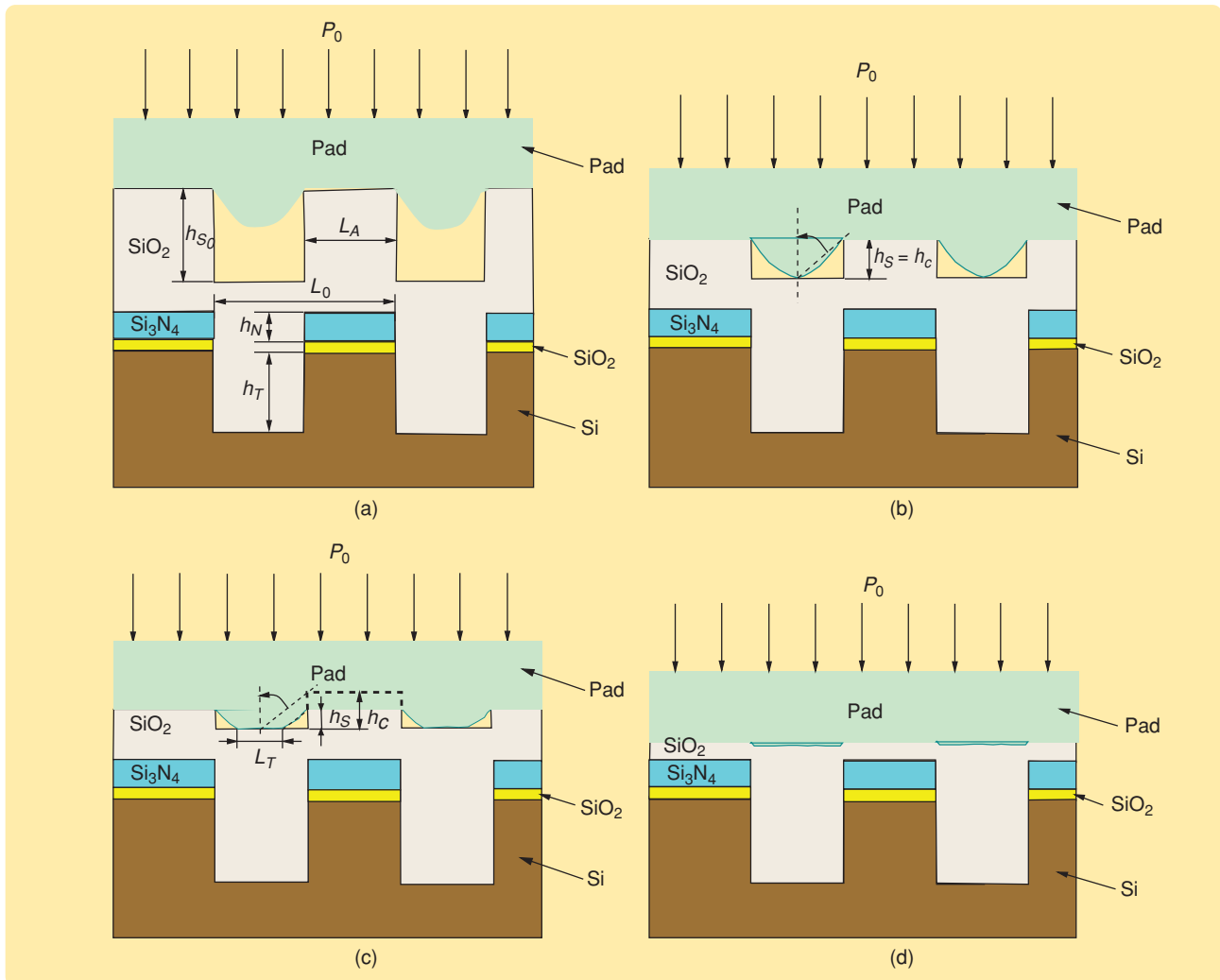
FIGURE 14 A snapshot of a cross-section of a shallow-trench-isolation patterned testing wafer during chemical-mechanical planarization. (a) The cross section at the beginning of the polishing process, when the initial step height h_{S_0} is larger than the contact height h_c, (b) the cross section when the pad starts contacting trenches, (c) the cross section when the pad touches the trenches, and (d) the patterned-wafer surface completely planarized. During the polishing process, the wafer/pad contact area is increasing, while the wafer/pad pressure is decreasing. (From [14].)

$$\gamma_{\text{patt}}(t) = \frac{1}{1 - \frac{h_{S_0}}{h_c} \rho e^{-\frac{t}{\tau_B}}}. \tag{18}$$

A detailed derivation of (17) and (18) is given in [14].

For the patterned STI wafer shown in Figure S2, the oxide film is deposited by a high-density plasma chemical vapor deposition process. The average thickness of the active oxide layer is 7053 Å, and the average trench depth is 7090 Å. The average thickness of the silicon nitride (Si_3Ni_4) layer is around 1503 Å. The step height h_{S_0} is around 5000 Å [15].

Since the density and pitch sizes of each subdie structure are different, as shown in Figure S2, to approximately estimate the pattern factor $\gamma_{\text{patt}}(t)$ for each die, we separate 25 subdie structures into two groups according to the two cases discussed above. In group I, $h_{S_0} < h_c$, while the group II structures have $h_{S_0} \geq h_c$. For the group I structures with pitch width $L_0 \geq 100\,\mu\text{m}$ and pattern density $\rho \geq 50\%$, an estimate of $h_c = 7000$ Å and $\tau_B = 20$ s is obtained. For group II, an estimate of $h_c = 4000$ Å is used [15]. Once the pattern factors $\gamma_{\text{patt}}^{\text{I}}$ and $\gamma_{\text{patt}}^{\text{II}}$ are obtained for groups I and II, respectively, an average pattern factor $\bar{\gamma}_{\text{patt}}$ for the whole die is finally calculated as

$$\bar{\gamma}_{\text{patt}} = \frac{n_{\text{I}}\gamma_{\text{patt}}^{\text{I}} + n_{\text{II}}\gamma_{\text{patt}}^{\text{II}}}{n_{\text{I}} + n_{\text{II}}}, \tag{19}$$

where $n_{\text{I}} = 12$ and $n_{\text{II}} = 13$ are the numbers of STI subdie structures in groups I and II, respectively. Figure 15 illustrates an estimate of $\bar{\gamma}_{\text{patt}}$, $\gamma_{\text{patt}}^{\text{I}}$, and $\gamma_{\text{patt}}^{\text{II}}$ for an STI test wafer.

EXPERIMENTAL MODEL VALIDATION

In this section, we first discuss the use of torque measurements to estimate model parameters and then present experimental model validation results.

We run experiments on a Lam Teres CMP polisher shown in Figure 16. A baseline oxide CMP is used to validate the friction models. Since the wafer/pad pressure distribution is determined mainly by the air-bearing system, a fixed configuration of the air-bearing system is used in all experiments to focus on the remaining process parameters. To start each experiment from the same pad surface condition, an identical ex situ pad-conditioning process is used in each run. Oxide wafers, Cabot SS12 slurry, and a Rodel IC 1000 pad are used in the experiments. The spindle torque is calculated based on the spindle mechanical components and

dc-motor parameters. The roller torque is captured by the roller motor current measurements. A detailed description of the hardware setup is given in [15].

The spindle friction torque M^s is used to estimate the model parameters μ_s/θ and σ. We rewrite (6) in the form

$$M^s = \mathbf{X}^T \Theta, \tag{20}$$

where

$$\mathbf{X}^T = \begin{bmatrix} \frac{p_0 \pi R^4 \omega}{4 v_B} & \frac{1}{2}\pi p_0 \omega R^4 \end{bmatrix} \tag{21}$$

and

$$\Theta = \begin{bmatrix} \frac{\mu_s}{\theta} & \sigma \end{bmatrix}^T. \tag{22}$$

A least-squares estimate $\hat{\Theta}$ of the model parameters in (20) is obtained from the experimental data. Several polishing processes with various low polishing pad speeds are designed. No pad conditioning is used in the experiments. Figure 17(a) illustrates the measured spindle torque at various pad speeds, shown as the solid curve, while the mechanical friction torque at the wafer rotational speed $\omega = 20$ rpm is shown as the dash-dot curve. The spindle
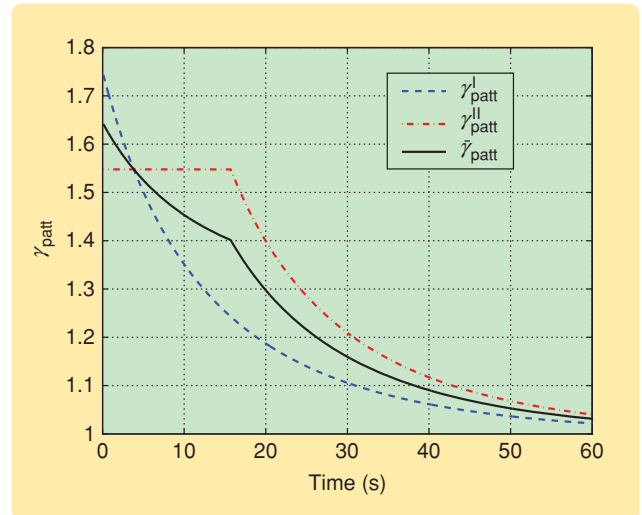
FIGURE 15 Estimated pattern factor γ_{patt} for polishing shallow-trench-isolation patterned testing wafers. The 25 subdie structures within one STI die are separated into two groups depending on their initial step heights. The pattern factors $\gamma_{\text{patt}}^{\text{I}}$ and $\gamma_{\text{patt}}^{\text{II}}$ are obtained for each group, after which the average pattern factor $\bar{\gamma}_{\text{patt}}$ for the entire die is calculated.

friction torque M^s is the difference between the solid and dash-dot curves in Figure 17(a). From the measured spindle torque M^s at $v_B = 50, 75, 100, 125$ ft/min, an estimate of the model parameters is obtained as

$$\hat{\Theta} = [0.323 \quad -0.085]^T. \tag{23}$$

Figure 17(b) shows the estimated and measured spindle torques as a function of pad speed v_B. The estimated spindle torque matches the measurements within 5%. We use the estimated model parameters in (23) to calculate the model predictions.

(a)

(b)

FIGURE 16 Lam Teres chemical-mechanical planarization (CMP) polisher, an integrated system for both oxide-wafer and copper-wafer planarization applications. (a) A 200-mm CMP polisher and (b) the top view of a 300-mm CMP polisher. An unprocessed wafer is picked up by a transfer robot from wafer cassettes and then passes through various process modules. The sequence through which a wafer undergoes various process modules is configurable by the process recipe. The polisher has two linear planarization process modules. After the polishing process, the wafer is cleaned and dried before it returns to the wafer cassettes. (Pictures used courtesy of Lam Research Corporation.)

We present two examples to validate the wafer/pad friction model. The first example, which is presented in this section, shows the oxide wafer polishing experiments with pad conditioning, while the second example demonstrates an STI patterned-wafer polishing process. The second example is also considered as an application example of the friction models and therefore is presented in the next section. A complete friction model validation is presented and discussed in [15].

Figure 18 indicates spindle and roller torques under various pad-conditioning settings, such as pad-conditioner down force and conditioning percentage. The conditioning percentage is defined as the ratio of the total conditioning process time to the total process time. We observe that the friction coefficient decays without pad conditioning, as the dashed curve shown in Figure 18(b). A similar observation is also obtained in [12]. This decaying trend matches the material removal rate drop reported in [8]. With pad conditioning, the friction torques are maintained at a constant level. Figure 18(b) shows that the roller torque is sensitive to the pad-conditioning variations. The higher conditioner down force increases the polishing pad roughness level, and thus yields a higher friction coefficient factor γ_μ in (11).

Figure 19 shows the measured and estimated spindle and roller friction torques for a baseline process with 50% pad conditioning. The model's predictions, shown as dotted curves, are within a 5% error range of the spindle and roller friction torque measurements, shown as solid curves, respectively. The spindle torque M^s is relatively small, which is consistent with (6) in the case of high pad speed v_B. For this reason, we do not use the spindle torque to estimate the decay of the wafer/pad friction coefficient. Instead, the roller-friction torque measurement is used to estimate the decay of μ_s in real time. A curve-fitting algorithm is used to estimate the decay of the friction coefficient μ_s in real time. The same curve is then used to predict the spindle torque, shown as the dotted curve in Figure 19(a). The estimated spindle curve matches the measured spindle torque, shown as the solid curve in the same figure, within a 5% error range.

APPLICATION EXAMPLES

We present three examples to illustrate the use of spindle and roller friction
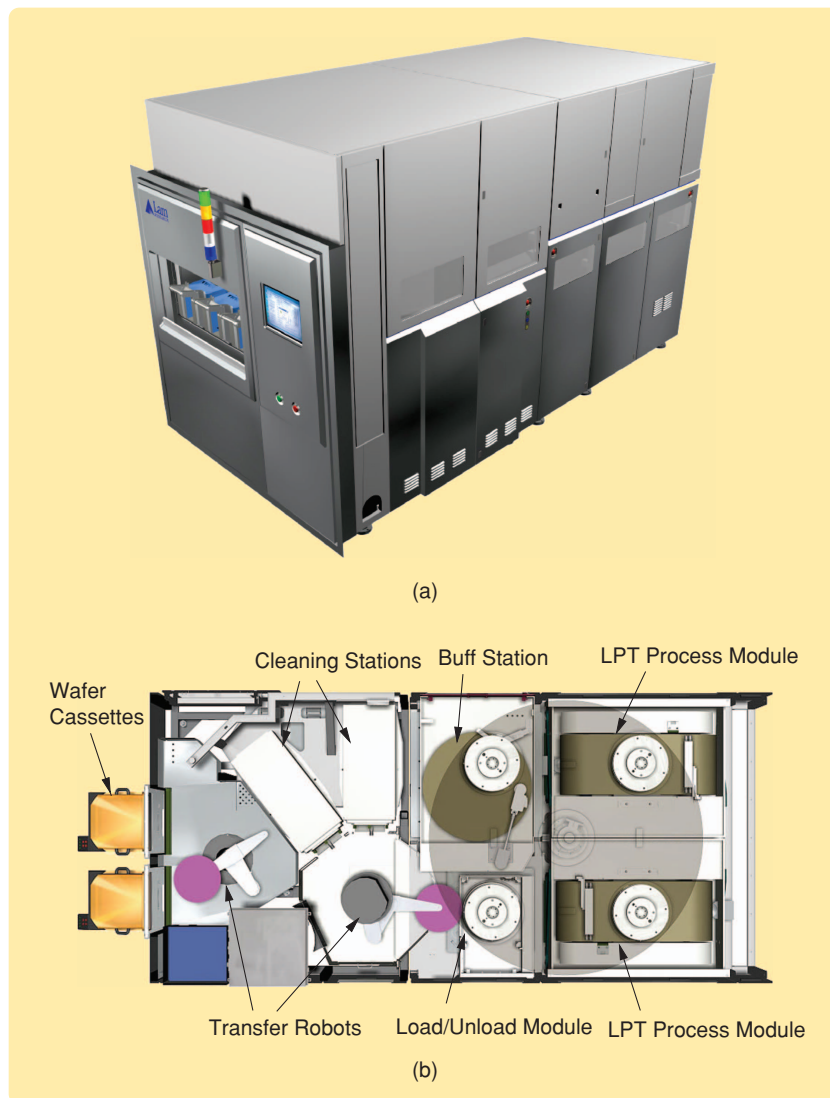
torque measurements to monitor the process performance. The first example of the STI patterned-wafer polishing process serves a dual role, namely, for friction-model validation and to predict the endpoint of the polishing process. We use the second and third examples to illustrate that the friction is an indicator of in situ CMP performance.

Example 1: STI CMP Endpoint Detection

Figure 20 shows the spindle and roller friction torque measurements during the polishing process of the STI test wafers. Torque measurements from polishing an oxide wafer using the same recipe are also plotted in Figure 20. Although no significant difference in spindle torques between the oxide and the STI patterned wafers is found, the roller torque measurements are different, as shown in

Figure 20(b). In particular, the large friction torque at the beginning of the STI polishing process is due to the film topography. As the process continues, the wafer surface is planarized (around 60 s) and the roller friction torque follows the oxide-wafer polishing trace. Finally, when the silicon nitride (Si_3Ni_4) layer is exposed, the wafer/pad friction increases due to a higher friction coefficient between the silicon nitride and the polishing pad.

An estimate of the roller torque measurement, shown as the dashed curve in Figure 20(b), is obtained by using the friction model and the data from polishing oxide wafers. The oxide wafer process data are available on a daily basis from wafer production. From Figure 20(b), we observe that the model-estimated roller torque trajectory fits the experimental data with a small range of error.
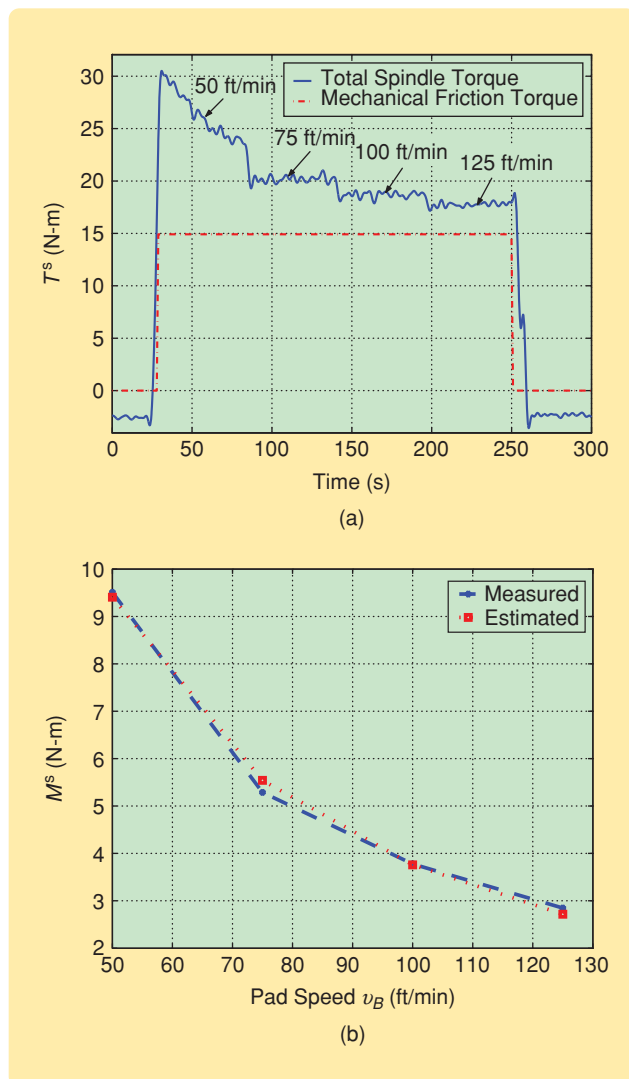
FIGURE 17 (a) The measured spindle torque at various pad speeds. The solid curve is the total driving spindle torque, while the dash-dot curve is the mechanical friction torque due to the rotational spindle mechanism. The difference between these curves is the wafer/pad friction torque M^s. (b) A comparison of the estimated and measured spindle torques. (From [15].)
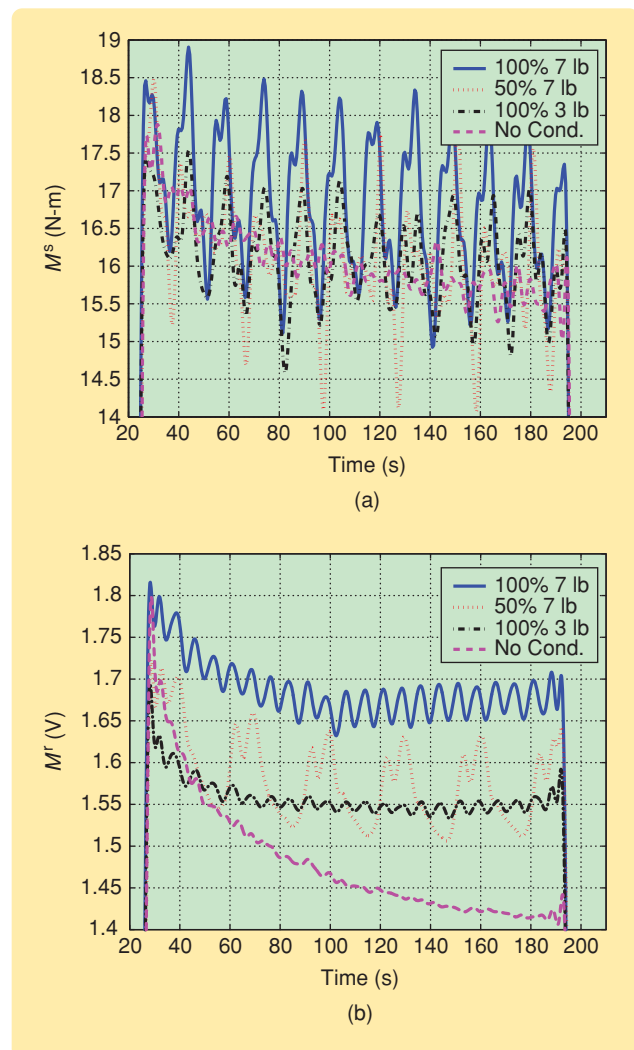
FIGURE 18 Comparison of various pad-conditioning experiments. (a) The polishing spindle torque and (b) the roller torque. Pad conditioning affects the wafer/pad friction significantly, as shown in (b). Without pad conditioning, the friction torque decreases, while with pad conditioning the friction is maintained at a stable level. A higher conditioner down force and a larger conditioning percentage also produce a higher friction torque. (From [15].)

The roller torque measurement in Figure 20(b) demonstrates the same trend as the material removal rate variation in the STI polishing process [25]. Moreover, from Figure 20(b), we observe an indication of the time instant (around 60 s) when the STI film is completely planarized, and the roller friction measurement can estimate the CMP endpoint. The roller torque measurements are used to predict the material removal rate and to determine the process endpoint in real time.

Example 2: 300-mm Copper CMP MRR Prediction

As an additional application, we consider 300-mm copper-patterned wafer removal rate monitoring. Figure 21 shows roller torque measurements for four 300-mm copper-patterned wafers, wafers #17, 77, 357, and 1014, from a large batch of polishing wafers in a nonstop marathon run over one week. The bulk copper CMP process is automatically stopped by an endpoint detection unit after a thickness of 5000 Å copper film is removed. A trend of increasing roller torque indicates faster material removal rate from wafers #17 to #1014. In particular, wafer #1014 has faster removal because a new pad conditioner disk is used right before that patterned wafer. Thus, the wafer/pad friction monitoring also indicates variations of the CMP process consumables. This observation is used for fault detection and diagnosis of polishing equipment and consumables [26].

Example 3: 300-mm Copper CMP Dishing Prediction

As a final example, we consider copper CMP dishing estimation. Figure 22 illustrates the relationship between the dishing average of 300-mm patterned wafers and the corresponding average roller friction torque measurements during a 700-wafer copper CMP run. The dishing average is calculated as the average dishing value of a total of 14 dies across the diameter of the 300-mm SEMATECH 754 patterned wafers. Dishing is a critical process specification for copper CMP, and it is difficult to obtain dishing measurements in real time. The square of the correlation coefficient, namely, $R^2 = 0.9529$ in Figure 22, demonstrates a statistically significant linear relationship between the dishing average and roller motor torque measurements. This linear relationship is used to monitor the dishing performance, prepare the consumables, and tune the process in real time [26].
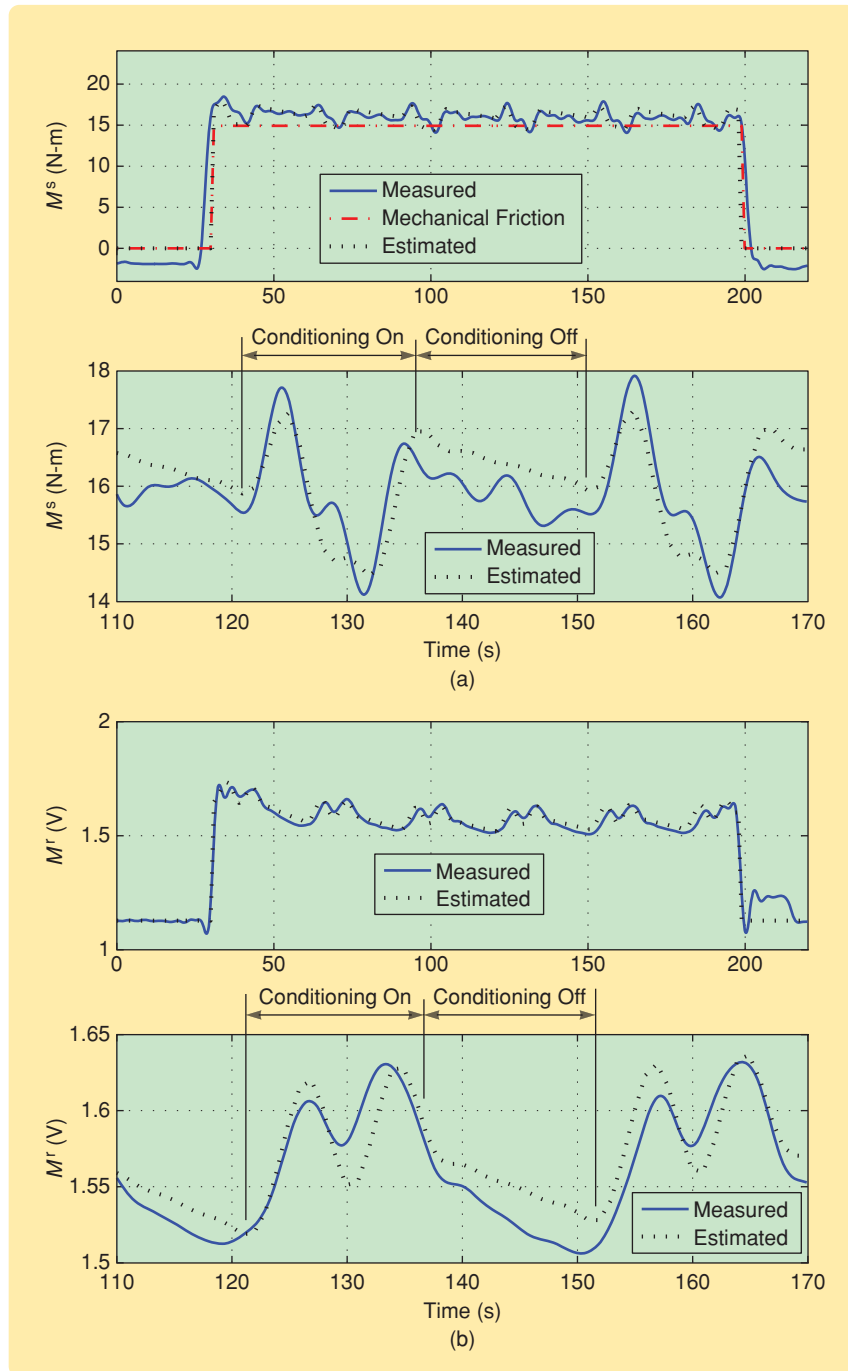
FIGURE 19 Model prediction and torque measurements for a baseline CMP process with 50% pad conditioning. (a) The estimated spindle friction torque, shown as the dotted curve in (a), follows the measurements, which are shown as the solid curve. (b) The estimated roller friction torque, shown as the dotted curve, follows the measurements, shown as the solid curve. The conditioning-on and -off periods during the polishing process are shown in both (a) and (b). (From [15].)

CONCLUSIONS

CMP is an enabling planarization technology in semiconductor manufacturing. Due to poor understanding of this complex process and a lack of in situ sensors, real-time monitoring and control of CMP is challenging. In this article, we report and demonstrate that wafer/pad friction modeling provides an effective way to monitor CMP in real time. We discuss and present an analytical friction modeling framework, experimental validation, and application examples for both oxide and copper CMP. A distributed LuGre dynamic friction model is used to capture the wafer/pad friction char-

acteristics. Since it is difficult to measure the wafer/pad friction directly, the polisher's spindle and roller-friction torque measurements are instead employed to monitor the friction without adding any new hardware. An analytical relationship between wafer/pad friction coefficients and polishing
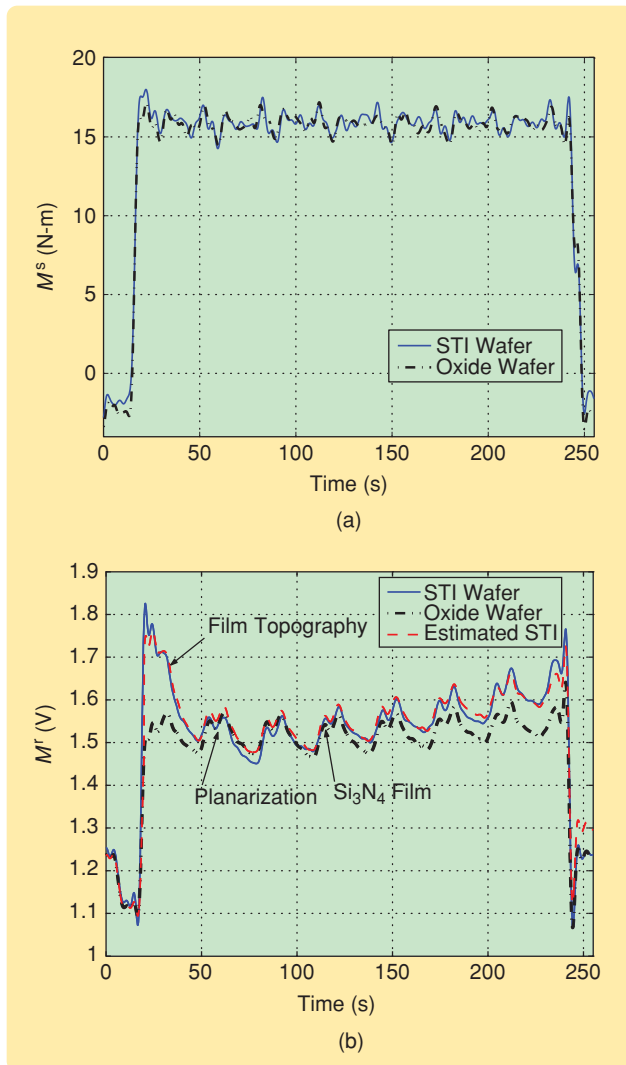

FIGURE 20 Comparison of polishing a shallow-trench-isolation patterned testing wafer, an oxide wafer, and model-based estimation. (a) The spindle friction torque shows no difference between polishing an STI wafer and an oxide wafer. (b) The roller-friction torque is higher than polishing the oxide wafer at the beginning of the polishing process due to the film topography. Once the wafer surface becomes planarized around 60 s, the roller-friction torque follows the oxide-wafer polishing trace (b). The roller-friction torque indicates the film topography evolution, and therefore predicts the polishing process endpoint. The friction model prediction also fits the experiments (b). (From [15].)
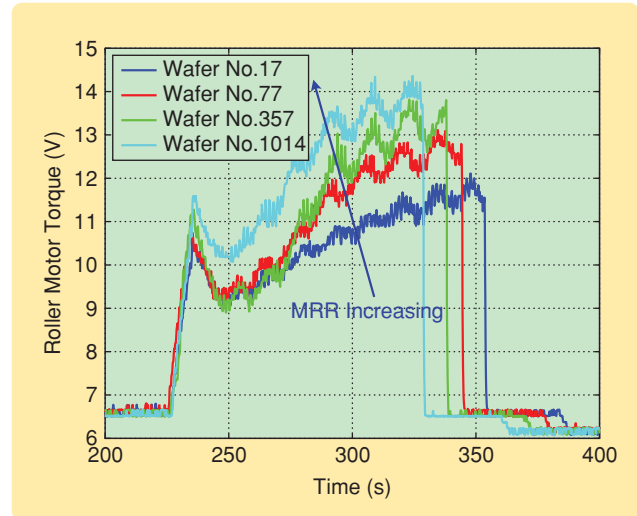
FIGURE 21 The roller-friction torque for polishing four selected 300-mm copper-patterned wafers from a non-stop copper-polishing process run over one week. The process stops automatically by an endpoint detection unit after a 5000-Å-thick copper layer is removed. Wafer #17 has the longest polishing time, and thus the slowest material removal rate, while wafer #1014 has the shortest processing time, and thus the fastest material removal rate. The roller-friction torque measurements show a trend of increasing material removal rate. The significant increase of material removal rate of wafer #1014 is due to the use of a new conditioner disk.
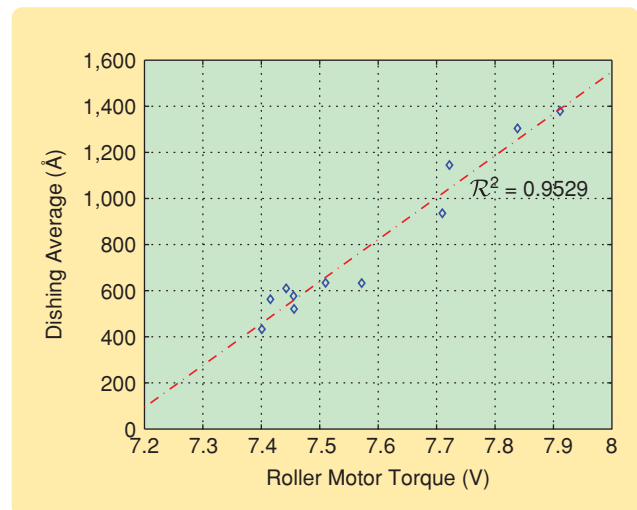
FIGURE 22 The relationship of the copper dishing average and the roller-friction torque for a set of Sematech patterned wafers during a 700-wafer copper-polishing process run. The dashed line is a linear fit. The correlation between the copper dishing average and the roller-friction torque is indicated by the square of the correlation coefficient, namely, $\mathcal{R}^2 = 0.9529$. The linear relationship between the dishing average and the roller-friction torque can be used for real-time monitoring of the copper-polishing process. (From [15].)

parameters is presented and discussed. The effects of pad conditioning and patterned-wafer topography on the wafer/pad friction characteristics are also investigated. The results show that the roller friction torque is a good candidate for monitoring wafer/pad friction. The models are used to in situ monitor process performance and further as a potential tool for polisher system fault diagnostics.

The friction modeling and analysis methods discussed in this article can be extended and applied to other types of CMP systems, such as rotary and orbital polishers. The modeling approaches derived in this article demonstrate that the knowledge and methodologies developed in friction modeling and control can be used to advance understanding and real-time monitoring of semiconductor manufacturing processes, such as CMP. The integration of the friction models into application examples are the future research. Another future research direction is to extend wafer/pad friction modeling to investigate how the additional process parameters, such as pad temperature and slurry chemical/mechanical characteristics, affect process performance.

ACKNOWLEDGMENTS

REFERENCES

[1] V. Martinez and T. Edgar, "Control of lithography in semiconductor manufacturing," *IEEE Control Syst. Mag.*, vol. 26, no. 6, pp. 46–55, 2006.

[2] J. Luo and D.A. Dornfeld, *Integrated Modeling of Chemical Mechanical Planarization for Sub-Micron IC Fabrication*. Berlin, Germany: Springer-Verlag, 2004.

[3] S. Limanond, J. Si, and K. Tsakalis, "Monitoring and control of semiconductor manufacturing processes," *IEEE Control Syst. Mag.*, vol. 18, no. 6, pp. 46–58, 1998.

[4] T. Edgar, S. Butler, W. Campbell, C. Pfeiffer, C. Bode, S. Hwang, K. Balakrishnan, and J. Hahn, "Automatic control in microelectronics manufacturing: Practices, challenges, and possibilites," *Automatica*, vol. 36, no. 11, pp. 1567–1603, 2000.

[5] J. Moyne, E. del Castillo, and A. Hurwitz, Eds., *Run-to-Run Control in Semiconductor Manufacturing*. Boca Raton, FL: CRC, 2001.

[6] J. Yi, Y. Sheng, and C. Xu, "Neural network based uniformity profile control of linear chemical-mechanical planarization," *IEEE Trans. Semiconduct. Manufact.*, vol. 16, no. 4, pp. 609–620, 2003.

[7] F. Preston, "The theory and design of plate glass polishing machines," *J. Soc. Glass Tech.*, vol. 11, pp. 214–256, 1927.

[8] H. Kim, H. Kim, H. Jeong, E. Lee, and Y. Shin, "Friction and thermal phenomena in chemical mechanical polishing," *J. Mat. Proc. Tech.*, vol. 130–131, pp. 334–338, 2002.

[9] D. DeNardis, J. Sorooshian, M. Habiro, C. Rogers, and A. Philipossian, "Tribology and removal rate characteristics of abrasive-free slurries for copper CMP applications," *Jpn. J. Appl. Phys.*, vol. 42, no. 11, pp. 6809–6814, 2003.

[10] Y. Homma, K. Fukushima, S. Kondo, and N. Sakuma, "Effects of mechanical parameters on CMP characteristics analyzed by two-dimensional frictional-force measurement," *J. Electrochem. Soc.*, vol. 150, no. 12, pp. G751–G757, 2003.

[11] Y. Homma, "Dynamical mechanism of chemical mechanical polishing analyzed to correct Preston's empirical model," *J. Electrochem. Soc.*, vol. 153, no. 6, pp. G587–G590, 2006.

[12] A. Sikder, F. Giglio, J. Wood, A. Kumar, and M. Anthony, "Optimization of tribological properties of silicon dioxide during the chemical mechanical planarization process," *J. Electr. Mater.*, vol. 30, no. 12, pp. 1520–1526, 2001.

[13] P. Carter and T. Werts, "A method for measuring frictional forces and shaft vibrations during chemical mechanical polishing," *J. Electrochem. Soc.*, vol. 154, no. 1, pp. H60–H66, 2007.

[14] J. Yi, "On the wafer/pad friction of chemical-mechanical planarization (CMP) processes, Part I: modeling and analysis," *IEEE Trans. Semiconduct. Manufact.*, vol. 18, no. 3, pp. 359–370, 2005.

[15] J. Yi, "On the wafer/pad friction of chemical-mechanical planarization (CMP) processes, Part II: Experiments and applications," *IEEE Trans. Semiconduct. Manufact.*, vol. 18, no. 3, pp. 371–383, 2005.

[16] A. Jensen, P. Renteln, S. Jew, C. Raeder, and P. Cheung, "Empirical-based modeling for control of CMP removal uniformity," *Solid State Technol.*, vol. 44, no. 6, pp. 101–106, 2001.

[17] T. Dyer and J. Schlueter, "Characterizing CMP pad conditioning using diamond abrasives," *Micro*, vol. 20, pp. 47–54, 2002.

[18] J. Luo and D. Dornfeld, "Material removal mechanism in chemical mechanical polishing: theory and modeling," *IEEE Trans. Semiconduct. Manufact.*, vol. 14, no. 2, pp. 112–133, 2001.

[19] J. Tichy, J. Levert, L. Shan, and S. Danyluk, "Contact mechanics and lubrication hydrodynamics of chemical mechanical polishing," *J. Electrochem. Soc.*, vol. 146, no. 4, pp. 1523–1528, 1999.

[20] C. Canudas de Wit, H. Olsson, K. Åström, and P. Lischinsky, "A new model for control of systems with friction," *IEEE Trans. Automat. Contr.*, vol. 40, no. 3, pp. 419–425, 1995.

[21] C. Canudas de Wit, P. Tsiotras, E. Velenis, M. Basset, and G. Gissinger, "Dynamic friction models for road/tire longitudinal interaction," *Veh. Syst. Dyn.*, vol. 39, no. 3, pp. 189–226, 2003.

[22] J. Yi, L. Alvarez, X. Claeys, and R. Horowitz, "Tire/road friction estimation and emergency braking control using a dynamic friction model," *Veh. Syst. Dyn.*, vol. 39, no. 2, pp. 81–97, 2003.

[23] J. Deur, J. Asgari, and D. Hrovat, "A 3D brush-type dynamic tire friction model," *Veh. Syst. Dyn.*, vol. 42, no. 3, pp. 133–173, 2004.

[24] T. Smith, S. Fang, D. Boning, G. Shinm, and J. Stefani, "A CMP model combining density and time dependencies," in *Proc. CMP-MIC Conf.*, Santa Clara, CA, 1999, pp. 97–104.

[25] S. Kim and Y. Seo, "Correlation analysis between pattern and non-pattern wafer for characterization of shallow trench isolation chemical-mechanical polishing (STI-CMP) process," *Microelectron. Eng.*, vol. 60, pp. 357–364, 2002.

[26] T. Travis, J. Yi, and P. Norton, "System and method for *in situ* characterization and maintenance of polishing pad smoothness in chemical-mechanical polishing," U.S. Patent 7,153,182, Dec. 2006.

[27] N.-H. Kim, Y.-J. Seo, and W.-S. Lee, "Temperature effects of pad conditioning process on oxide CMP: Polishing pad, slurry characteristics, and surface reactions," *Microelectron. Eng.*, vol. 83, pp. 362–370, 2006.

AUTHOR INFORMATION

Jingang Yi (jgyi@jove.rutgers.edu) received the B.S. in electrical engineering from the Zhejiang University, China, in 1993, the M.Eng. in precision instruments from Tsinghua University, China, in 1996, the M.A. in mathematics, and the Ph.D. in mechanical engineering from the University of California, Berkeley, in 2001 and 2002, respectively. He is an assistant professor in mechanical and aerospace engineering at Rutgers University. His research interests include autonomous robotic systems, dynamic systems and control, intelligent sensing and actuation systems, and automation science and engineering, with applications to semiconductor manufacturing and intelligent transportation systems. He is a member of the American Society of Mechanical Engineers (ASME) and a Senior Member of the IEEE. He can be contacted at the Department of Mechanical and Aerospace Engineering, Rutgers University, 98 Brett Road, Piscataway, NJ 08854-8058 USA.

The Z-Properties Chart

BRIAN S.R. ARMSTRONG and QUNYI CHEN

VISUALIZING THE PRESLIDING BEHAVIOR OF STATE-VARIABLE FRICTION MODELS

I n recent years, dynamic friction modeling has received considerable attention, and diverse models have been put forward [1]–[9]. Some of the most widely applied models are state-variable models, which incorporate one or more internal states governed by nonlinear differential equations. As empirically motivated nonlinear differential equations, state-variable friction models exhibit a surprising variety of behavior. Most of these models are dissipative for motions on all scales, while others are nondissipative for very small motions; some show true sliding for very small applied forces, while others show purely elastic displacement under the same conditions; and some show hysteresis with nonlocal memory, while others do not.

In this article the Z-properties chart is introduced. This chart provides a means for comparing and contrasting the properties of the state-variable friction models under the condition referred to as presliding, where the dynamics of the state variables play their greatest role. The friction state is often designated $z(t)$ and thus the name "Z-properties chart."

MICROSCALE FEATURES OF FRICTION

Early developments in friction modeling, such as Tustin's work on frictional limit cycles, employ the Coulomb friction model [10]. This model, illustrated in Figure 1, represents friction solely as a function of velocity $v(t)$ and has a discontinuity at $v(t) = 0$. An illustrative velocity profile with two zero crossings is shown in Figure 2(a), with the friction force given by the Coulomb model in Figure 2(b).

Experiments with servo control and micron-scale precision reveal a friction-time profile during velocity reversal that is more complex than that of the Coulomb friction model [2], [11]. The detailed dynamics of friction are beyond the scope of this article. For surveys addressing frictional dynamics and control, see [12] and [13]. When

high-precision measurements are made, an observed feature of the frictional dynamics is a smooth transition in friction force during velocity reversal, as illustrated in Figure 3. Working with a precision pointing system, Philip Dahl of The Aerospace Corporation considered what is now known as the Dahl friction model, given by

$$\dot{z}(t) = \left(1 - \text{sgn}(v(t))\frac{z(t)}{z_{\text{ss}}}\right)^i v(t), \tag{1}$$

$$f_{\text{f}}(t) = \sigma_0 z(t), \tag{2}$$

where $v(t)$ is the velocity of the sliding body, $f_{\text{f}}(t)$ is the friction force, σ_0 is a stiffness parameter, z_{ss} is a parameter specifying the maximum excursion of $z(t)$, and $z(t)$ is an internal state of the friction model [1], [2]. Additionally,
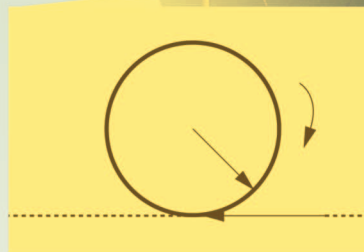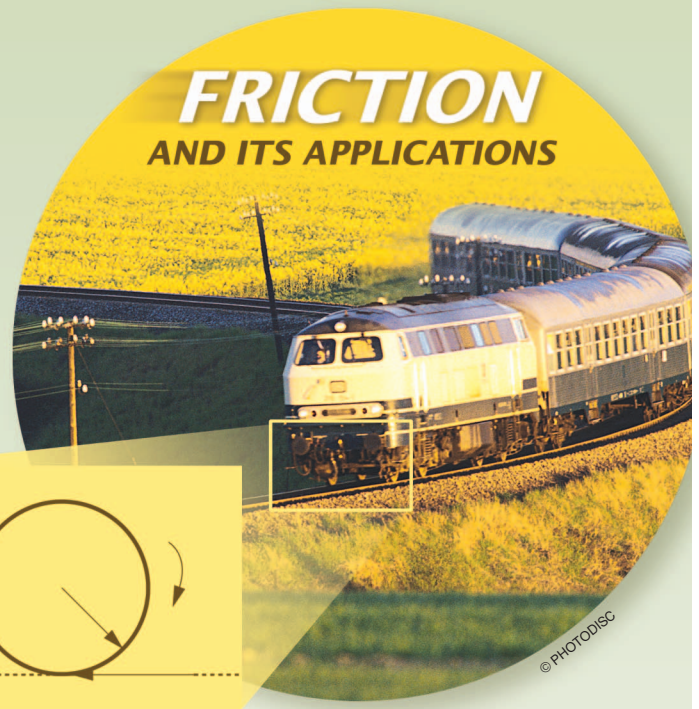
FIGURE 1 The Coulomb friction model relating velocity $v(t)$ to friction force $f_f(t)$. Static friction is shown at zero velocity. This friction model, while adequate for many purposes, does not accurately model friction phenomena during velocity reversals.

the parameter i is used to tune the shape of the friction-displacement curve during presliding displacement [2], [14].
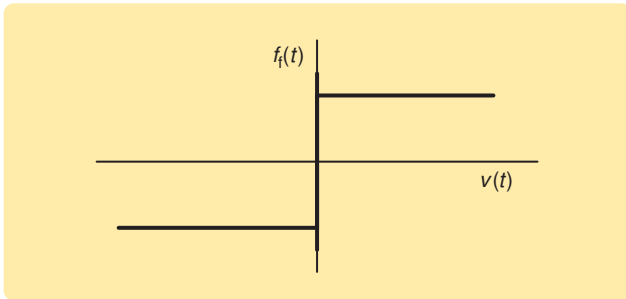
The model (1), (2) has the property that the friction transition during velocity reversal is governed by *displacement* rather than velocity. In fact, a time-free formulation of the Dahl model can be obtained by multiplying both sides of (1) by dt, namely,

$$dz(t) = \left(1 - \text{sgn}(v(t))\frac{z(t)}{z_{ss}}\right)^i dx(t), \qquad (3)$$

where $x(t)$ is the position at time t. By integrating both sides of (3) over an interval without velocity reversal, $z(t)$ is found as a function of $x(t)$, with no explicit dependence on time or velocity [15]. Models such as (1), with one or several internal states, are referred to as state-variable friction models.
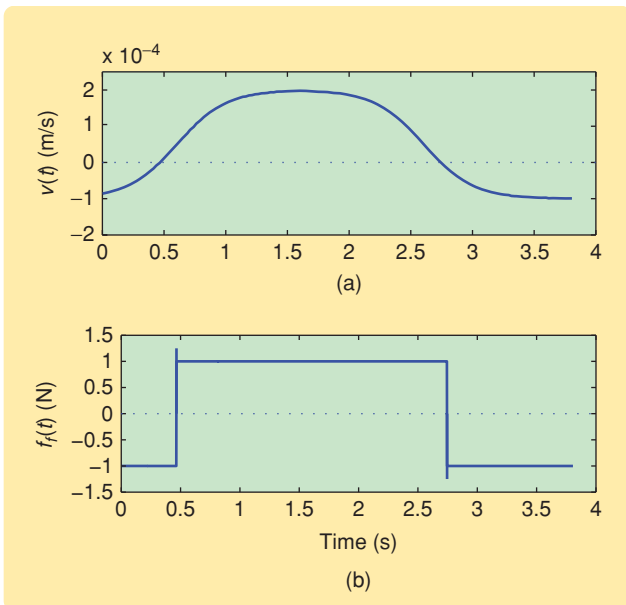

FIGURE 2 Illustrative velocity profile and friction force predicted by the Coulomb friction model. (a) A velocity-time profile and (b) the corresponding friction-time profile. The velocity profile crosses zero at $t = 0.5$ s and $t = 2.75$ s. The Coulomb friction model predicts an instantaneous transition between negative and positive friction force when the velocity crosses zero.

On the microscopic scale, the interaction between sliding surfaces is complex. Figure 4 illustrates some of the contributing phenomena. The rubbing surfaces of bearings, gears, and other parts that make up machines are rough on a microscopic scale. Contact occurs over only a small fraction of the total surface area, between surface features known as *asperities*. Qualitatively, the internal state of the friction model can be imagined as being associated with the elasticity of the asperities.

A friction interface exists between two bodies in contact with relative motion. In state-variable friction models, the externally measurable displacement between the bodies is comprised of two components

$$x(t) = z(t) + w(t), \qquad (4)$$

where $x(t)$ is the externally measurable displacement, $z(t)$ is the state variable of the friction model and reflects elastic displacement, and $w(t)$ is the true sliding of one body past the other. These terms are illustrated schematically in Figure 5, which shows a mass sliding with friction. Because the elastic displacement occurs within the friction junction, neither $z(t)$ nor $w(t)$ can be directly sensed.

Several terms are helpful for discussing the possible motions of a state-variable friction model. Considering Figure 5, *true sliding* refers to displacement of the asperities of one body relative to the other, or $\dot{w}(t) \neq 0$. *Fully developed sliding* refers to the condition where $z(t)$ has reached its maximum extension. Fully developed sliding exists when $v(t) \neq 0$ and $\dot{z}(t) = 0$. For example, in (1) fully developed sliding exists when $v(t) > 0$ and $z(t) = z_{ss}$ or $v(t) < 0$ and $z(t) = -z_{ss}$. *Presliding displacement* refers to motions that include a component of elastic displacement, that is, $v(t) \neq 0$ and $\dot{z}(t) \neq 0$. Finally, *steady sliding* refers to sliding with constant velocity.

An analogy is drawn between the characteristics of state-variable friction models and elastoplastic displacement in mechanics [7]. In this analogy, the friction-model state variable $z(t)$ corresponds to elastic displacement, while the true
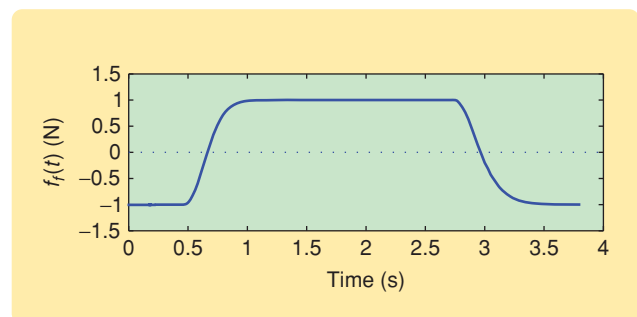

FIGURE 3 A smooth friction-time curve, as predicted by the Dahl friction model applied to the velocity profile of Figure 2(a). This curve can be contrasted with the friction-time profile of Figure 2(b). The Dahl friction model, which has an internal state $z(t)$, predicts a friction-time profile without discontinuity.

sliding $w(t)$ corresponds to plastic deformation. The elastic characteristic of $z(t)$ is seen in (2), where the friction force depends on $z(t)$ through the stiffness parameter σ_0. The state variable of each state-variable friction model has the behavior of elastic displacement. The true sliding $w(t)$ of one body past the other is analogous to plastic displacement in elastoplastic mechanics in that both give rise to hysteresis, and both are associated with energy dissipation. The explicit consideration of $w(t)$, in addition to the externally measurable displacement $x(t)$ and the friction-model elastic state $z(t)$, reveals properties of the state-variable friction models. Part of the challenge of friction modeling lies in the fact that only $x(t)$ is externally observable. The decomposition of $x(t)$ into $z(t)$ and $w(t)$ occurs within the sliding contact and can only be inferred from the externally observable force $f_f(t)$ in (2) and motions $x(t)$ and $v(t)$ in (1), (3), and (4).

When a friction model has a single state variable, the state can be thought of qualitatively as representing the deflection of an asperity or bristle in the friction interface [3]. Modeling the bristle as having stiffness σ_0 yields (2) for the friction force. In lubricated steel-on-steel contacts, the maximum deflection parameter z_{ss} in (1) may be on the order of 5 (μm), while the stiffness σ_0 in (2) may range from thousands to millions of newtons per meter, depending on the friction level [12].

State-variable friction models are empirical. Some of these models can give accurate predictions of measured friction [11] and lead to improved precision control [16]. Additionally, by eliminating the discontinuity at $v(t) = 0$ from the friction model, state-variable friction models are well suited to numerical simulation. The difficulties of numerical simulation that can arise with a discontinuous friction model such as the Coulomb friction model are addressed in [17]–[19]. State-variable friction models avoid the discontinuity and can thus be accommodated by standard numerical integrators. However, state-variable models often require a numerical integrator with an adaptive step size, since they may be extremely stiff near velocity zero crossings. State-variable models are widely used in the control literature [1]–[3], [5]–[9], [20].

In the following section we consider seven state-variable friction models, namely, the integrator reset model [3], the Dahl model [1], [2], the LuGre model [5], the elastoplastic model [7], the Leuven and modified Leuven models [6], [8], the Maxwell-slip (MS) model [21], [22], and the generalized generalized Maxwell-slip (G-GMS) model [9], [20]. A shared feature of these state-variable models is that each has an internal state or states with units of length or angle, which qualitatively represents elastic deformation in the
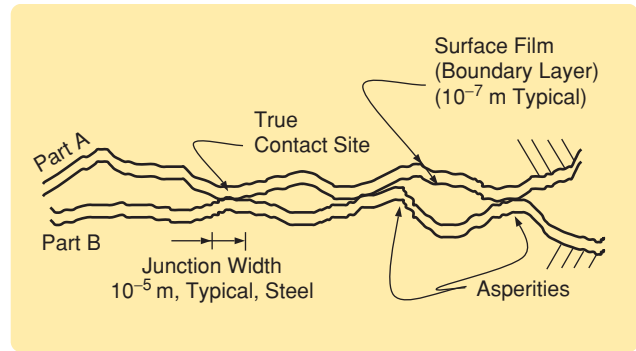

FIGURE 4 Schematic illustration of a rubbing contact, particularly a lubricated metal-on-metal contact of the type often found in servo machines. Many surfaces are irregular on a microscopic scale, with elevated features called asperities. Contacts that occur between asperities may constitute a small fraction of the total surface area. Lubricants or oxidation can provide a boundary layer with complex chemical and mechanical properties. The boundary layer can play a dominant role in the resulting friction.

frictional interface. Additionally, each state-variable model has a nonlinear differential equation governing the evolution of the friction state or states. For consistency, each model is written using the notation introduced with the LuGre model, that is, $z(t)$ as the state variable and σ_i as the friction parameters. Although linear sliding coordinates are used throughout, the motion and state variables can also be expressed in rotational coordinates.

Systems governed by nonlinear differential equations can exhibit a remarkable wealth of behaviors. Even though the considered models have similar governing differential equations, they can have quite different macroscopic behavior. The Z-properties chart is a tool for charting, comparing, and contrasting the behavior of state-variable friction models.

THE Z-PROPERTIES CHART
The insight of the Z-properties chart is that, for the state-variable friction models, several dynamic-model
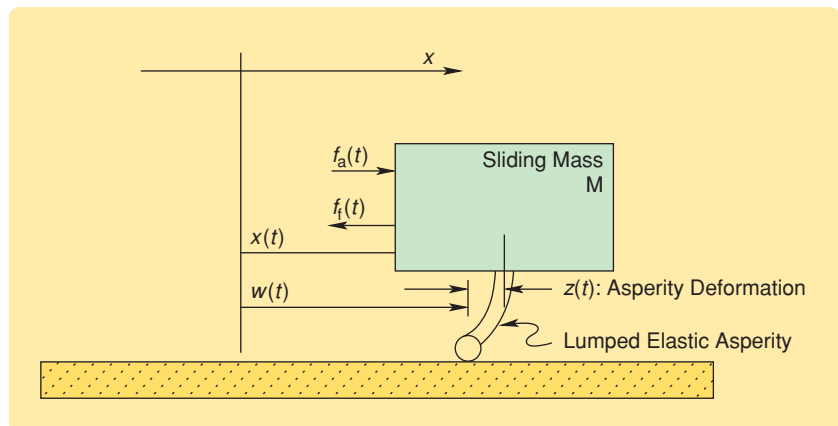

FIGURE 5 Schematic of frictional contact showing externally measurable displacement $x(t)$ broken down into elastic displacement $z(t)$ and true sliding $w(t)$. The term $f_a(t)$ is the applied force, and $f_f(t)$ is the friction force.

properties are present or absent based solely on the magnitude of the state variable. In the Z-properties chart, the vertical axis corresponds to $z(t)$, and the regions of $z(t)$ on which various model properties hold are indicated.

For example, if a hypothetical state-variable friction model shows properties A, B, and C, then the Z-properties chart might look like Figure 6, where property A is exhibited by the model for all values of $z(t)$, property B is shown for $z_2 \leq z(t)$, and property C is exhibited on the range $z_2 \leq z(t) \leq z_{ss}$.

As seen in Figure 6, only positive values of $z(t)$ are considered in the Z-properties chart, since the models consid-
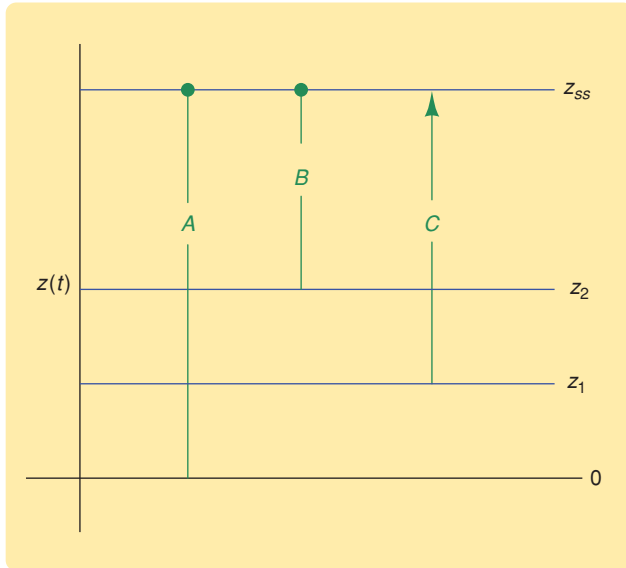

FIGURE 6 Illustrative Z-properties chart. The vertical axis qualitatively indicates the level of $z(t)$. The symbols A, B, and C indicate properties, while the levels z_1, z_2, and z_{ss} indicate magnitudes of $z(t)$ at which the model properties change. The case $z(t) < z_{ss}$ corresponds to the presliding or microdynamic regime, whereas $z(t) = z_{ss}$ corresponds to steady or continuous sliding. Properties A and B are exhibited in the presliding regime and in steady sliding (shown with a dot at $z(t) = z_{ss}$). Property C is not exhibited in steady sliding. The qualitative friction model properties are identical for positive- and negative-direction sliding, and thus only positive values of $z(t)$ are shown in the Z-properties charts.

TABLE 1 State-variable friction model properties with labels used in the Z-properties charts. For example, a region in the Z-properties chart marked iH indicates rate-independent hysteresis for values of $z(t)$ within the region.

Property	Label
Rate-independent hysteresis	iH
Rate-dependent hysteresis	dH
Hysteresis with nonlocal memory	N
Damping of the microdynamics	d
Model can exhibit drift	D
Drift is precluded	$\overline{D}$
True sliding (sliding with $\dot{z}(t) \neq \dot{x}(t)$)	tS

ered have the same properties for positive- and negative-direction sliding. The levels z_1, z_2, and z_{ss} at which model properties change are determined by the equations of each model, as discussed below. A mathematical property of (1), as well as all state-variable friction models, is that $|z(t)|$ converges to z_{ss} during steady sliding. In Figure 6, properties A and B are exhibited in sliding, as indicated by a ball on the $z(t) = z_{ss}$ line, whereas property C terminates with the transition to steady sliding.

We use the Z-properties chart to exhibit six properties, which are enumerated in Table 1. Some models show rate-independent or rate-dependent hysteresis, which may additionally have the property of nonlocal memory. In this article, only hysteresis in the presliding regime is considered. All friction models show hysteresis for fully developed sliding.

The state $z(t)$ of a state-variable friction model is associated with energy storage. Because the frictional interface is often very stiff, such as with $\sigma_0 = 10^5$ N/m, the range of motion for $z(t)$ is generally quite small. The dynamics involving the response of the friction state constitute the *microdynamics* of the system [23], whereas the dynamics corresponding to fully developed sliding are referred to as the *macrodynamics*. A mechanism for dissipating friction-state energy is incorporated in several models to damp the microdynamics, as indicated by d in the Z-properties charts.

Several of the state-variable friction models exhibit drift. In this context, drift is net displacement that occurs due to an arbitrarily small level of applied force, typically including an oscillatory component. Drift is illustrated in Figure 7(b), which shows the LuGre friction model [5] simulated with a Coulomb friction level $F_c = 1.0$ N and contact stiffness $\sigma_0 = 10^5$ N/m. The elastoplastic friction model, illustrated in Figure 7(c), does not exhibit drift under the illustrated test conditions, although it does exhibit elastic displacement. Regimes on which drift is possible are labeled D in the Z-properties charts, whereas regimes on which drift is precluded are labeled $\overline{D}$. Finally, the regime with a component of true sliding, that is, where $\dot{w}(t) \neq 0$, is indicated by tS in the Z-properties charts.

THE INTEGRATOR-RESET MODEL

The state equation for the integrator reset model is given by

$$
\dot{z}(t) = \begin{cases} 0, & v(t) > 0 \text{ and } z(t) \geq z_{ss}, \\ 0, & v(t) < 0 \text{ and } z(t) \leq -z_{ss}, \\ v(t), & \text{otherwise,} \end{cases} \tag{5}
$$

and the associated output equation is given as

$$
f_f(t) = \begin{cases} \sigma_0 z(t) + \sigma_1 \dot{z}(t), & |z(t)| < z_{ss}, \\ \sigma(v(t)) z_{ss} \operatorname{sgn}(z(t)), & \text{otherwise,} \end{cases} \tag{6}
$$

where σ_0 is a stiffness constant, σ_1 is a damping parameter, $z_{ss} = f_C/\sigma_0$ is the maximum value for $z(t)$, f_C is the

Coulomb friction level, and $v(t) = \dot{x}(t)$ is the externally measurable velocity across the frictional contact [3]. During sliding, the stiffness constant σ_0 is replaced by a function of velocity $\sigma(v(t))$ to represent the velocity dependence of friction [3].

The Z-properties chart for the integrator reset model given in Figure 8 shows that this model has damped microdynamics and is drift free for all values $|z(t)| < z_{ss}$. With the double horizontal lines, the chart shows that the model changes structure at $|z(t)| = z_{ss}$, corresponding to the transition from case 1 or case 2 to case 3 in (5). Additionally, the absence of ranges labeled iH and dH indicates that the integrator reset model does not show hysteresis for any value of $z(t)$ in the presliding regime.

THE DAHL MODEL

For the Dahl model given by (1), (2), the Z-properties chart is shown in Figure 9. Unlike the integrator reset model, the Dahl model shows true sliding, rate-independent hysteresis and drift for the entire range of $z(t)$, while no change of model structure occurs when $|z(t)| = z_{ss}$.

THE LUGRE MODEL

The LuGre model, which is named for the cities of Lund and Grenoble [5], demonstrates that a state-variable friction model can reflect several dynamic friction characteristics with a single model structure, particularly the nonlinear friction-velocity curve known as the Stribeck curve [24], [25]. The LuGre model can be written as

$$\dot{z}(t) = \left(1 - \frac{z(t)}{z_{ss}(v(t))}\right) v(t), \tag{7}$$

$$f_f(t) = \sigma_0 z(t) + \sigma_1 \dot{z}(t) + \sigma_2 v(t), \tag{8}$$

where $z_{ss}(v(t)) = f_{ss}(v(t))/\sigma_0$ is the steady-state value of $z(t)$ with steady sliding at velocity $v(t)$, $f_{ss}(v)$ captures the friction-velocity curve for steady sliding at velocity $v(t)$, and the nonnegative parameters σ_1, σ_2 add damping.

The Z-properties chart for the LuGre model in Figure 10 is similar to that of the Dahl model shown in Figure 9. Like the Dahl model, the LuGre model can show drift for small applied forces [7] and does not show hysteresis with non-local memory [6], [26]. When a nonlinear friction-velocity relationship is modeled with $f_{ss}(v)$, the LuGre model shows rate-dependent hysteresis.
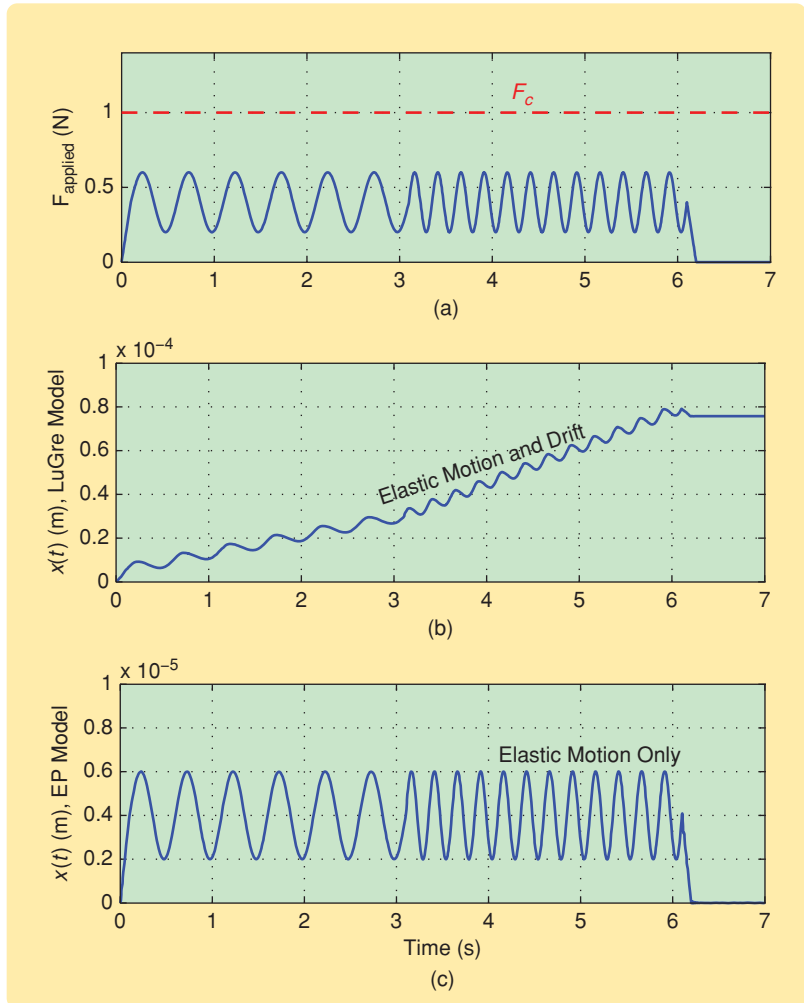
FIGURE 7 Simulation of state-variable friction models with and without drift. (a) The simulated applied force, which comprises a bias plus oscillatory component and never exceeds the Coulomb friction level. In (b) the externally measurable response $x(t)$ of the LuGre model exhibits steady net motion or drift, while in (c) the externally measurable response of the elastoplastic model exhibits elastic deformation but no drift.
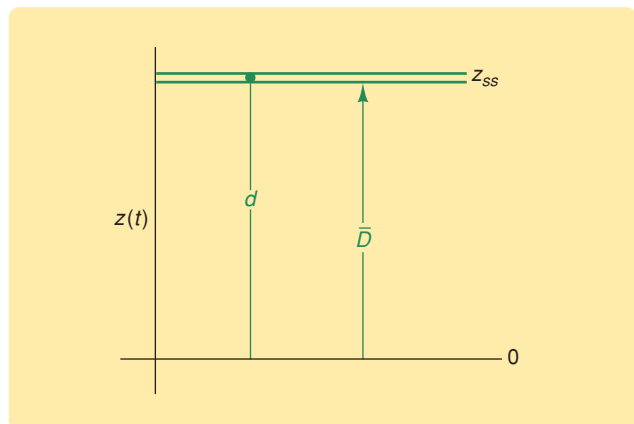
FIGURE 8 Z-properties chart for the integrator reset model. Over the range $z(t) < z_{ss}$, the model exhibits damped microdynamics (d) and no drift ($\overline{D}$). The double heavy line at $z(t) = z_{ss}$ indicates a change in the structure when $z(t) = z_{ss}$.

FIGURE 9 Z-properties chart for the Dahl model. Over the entire range $z(t) \le z_{ss}$, the model exhibits damped microdynamics (d) and true sliding (tS). On the range $z(t) < z_{ss}$, the model exhibits rate-independent hysteresis (iH) and drift (D).

FIGURE 10 Z-properties chart for the LuGre model. Like the Dahl model, the LuGre model exhibits damped microdynamics (d) and true sliding (tS) over the f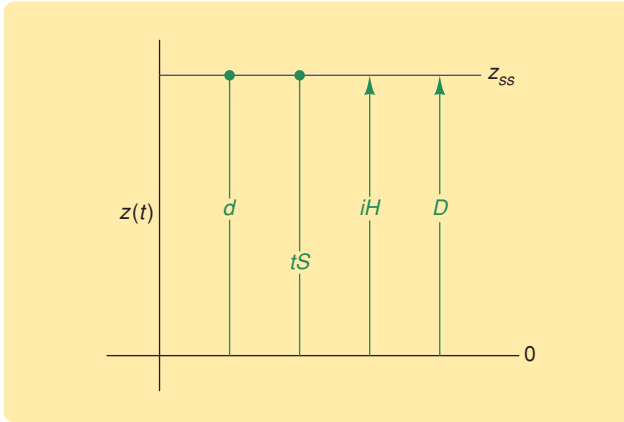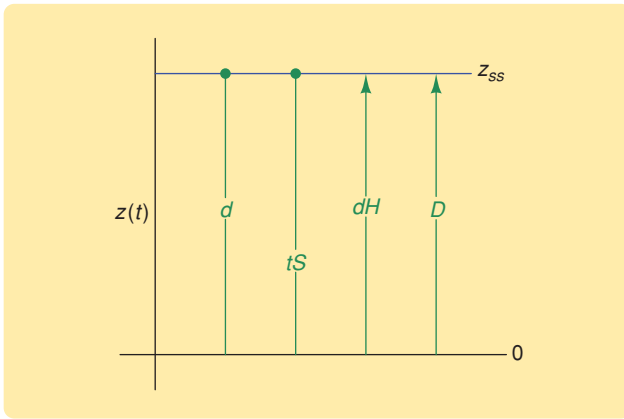ull range of $z(t)$, along with drift (D) when $z(t) < z_{ss}$. Because of the term that captures the Stribeck curve, the model exhibits rate-dependent hysteresis (dH).
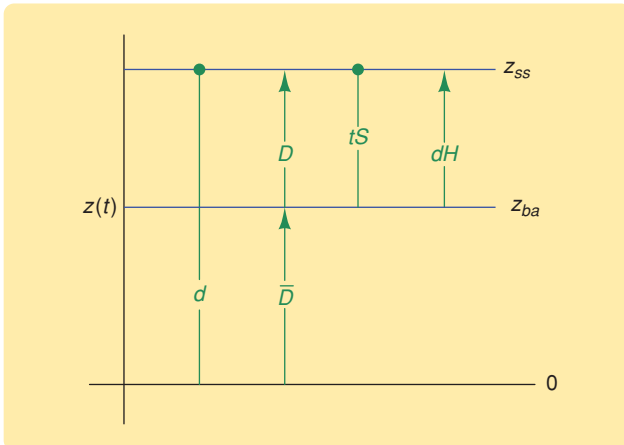

FIGURE 11 Z-properties chart for the elastoplastic friction model. At $z(t) = z_{ba}$ the model has a transition. For $z(t) \le z_{ba}$, the model does not drift ($\overline{D}$), whereas, for $z_{ba} < z(t) \le z_{ss}$, the model exhibits drift (D), true sliding (tS), and rate-dependent hysteresis (dH). The model has dissipative microdynamics (d) for the entire range of $z(t)$.

THE ELASTOPLASTIC MODEL

The elastoplastic friction model is given by

$$\dot{z}(t) = \left(1 - \alpha(z, v)\frac{z(t)}{z_{ss}(v(t))}\right) v(t), \tag{9}$$

$$f_f(t) = \sigma_0 z(t) + \sigma_1 \dot{z}(t) + \sigma_2 v(t), \tag{10}$$

where $\alpha(z, v)$ is zero when $\text{sgn}(z) \ne \text{sgn}(v)$ and $|z(t)| \le z_{ba}$, where z_{ba} is a parameter of the model [7]. By setting $\alpha(z, v) = 0$, a purely elastic sliding range is created, as in the integrator reset model. The Z-properties chart for the model shown in Figure 11 is distinct from the charts of figures 9 and 10 in that a new level z_{ba} is introduced. When $|z(t)| < z_{ba}$, the displacement is purely elastic and no drift occurs, whereas, for $|z(t)| \ge z_{ba}$, plastic displacement and hysteresis arise, but not hysteresis with nonlocal memory.

THE LEUVEN MODEL

When a process with hysteresis undergoes velocity reversals, it is possible for a *minor hysteresis loop* to be created, which is contained entirely within an enclosing *major hysteresis loop* [6], [26]–[28]. An example is seen in Figure 12, which illustrates friction as a function of displacement for an electrical discharge machine tool. For a hysteresis process with nonlocal memory, the friction-displacement curve returns to the trajectory of the enclosing major hysteresis loop when the minor loop is closed.

The Leuven model combines a Dahl-like state equation with two specialized data structures, called *hysteresis stacks*, which capture the friction-model state at each instant of velocity reversal [6]. The hysteresis stacks are implemented in computer memory as part of the Leuven model and operate in the same way as a microprocessor stack. In the operation of a microprocessor, data are *pushed* onto the stack when entering a function and *popped* off the stack when the function is completed. The stack is referred to as a last-in, first-out (LIFO) memory, since the most recently stored data is the first data recovered. In the Leuven friction model, the friction state is pushed onto the stack at the instant of a velocity zero crossing and popped off the stack when a hysteresis loop is closed.

The state equation of the Leuven model is given by

$$\dot{z}(t) = \left(1 - \text{sgn}\left(\frac{f_d(z)}{f_{ss}(v) - f_b(t)}\right)\left|\frac{f_d(z)}{f_{ss}(v) - f_b(t)}\right|^n\right) v(t), \tag{11}$$

where $v(t)$ is the velocity, $f_{ss}(v)$ is the friction-velocity curve for steady sliding, n is empirically tuned, and $f_d(z(t))$ is a "point-symmetrical strictly increasing function of $z(t)$" [6], the simplest of which corresponds to a linear

spring given by $f_d(z) = \sigma_0 z$. The value $n = 7$ is used in [6]. The friction force is given as

$$f_f(t) = f_h(t) + \sigma_1 \dot{z}(t) + \sigma_2 v(t), \qquad (12)$$

with $f_h(t)$ given by

$$f_h(t) = f_b(t) + f_d(z(t)), \qquad (13)$$

and where $f_b(t)$ is set by the rules of the hysteresis stack mechanism. As described in [6], the hysteresis stacks operate according to the following three rules.

Rule 1

Record $f_h(t)$ at each instant of velocity reversal. At a transition from negative to positive velocity, the current value of $f_h(t)$ is recorded to the hysteresis stack m, whereas at a transition from positive to negative velocity, the current value of $f_h(t)$ is recorded to hysteresis stack M. Additionally, at each velocity reversal $f_b(t)$ is set equal to the current value of $f_h(t)$, and $z(t)$ is set to zero. Thus, $f_h(t)$ is a continuous function of time, but $f_b(t)$, $f_d(z)$, and $z(t)$ are adjusted in a discontinuous way. Let $n_m(t)$ and $n_M(t)$ be the number of values on hysteresis stacks m and M, respectively. When $n_m \geq 2$ or $n_M \geq 2$, motion has reversed at least twice, and the system is in a minor hysteresis loop.

Rule 2

When $\dot{x} > 0$ and $f_h(t) = M(n_M(t))$, or when $\dot{x} < 0$ and $f_h(t) = m(n_m(t))$, the system leaves the current minor hysteresis loop. An example is annotated "minor loop closed" in Figure 12. Using the information stored on the hysteresis stack, f_b, $z(t)$, $f_d(z(t))$ and $f_h(t)$ are restored to their values on the enclosing hysteresis loop. For example, if $\dot{x} > 0$ and $f_h(t)$ equals the top value on stack M, the values are restored to those that existed at the instant the top value on stack M was recorded.

Rule 3

When $\dot{z}(t)$ becomes zero while $v(t) \neq 0$, the system enters fully developed sliding, and the hysteresis stacks are reset, all values on the stacks are cleared.

The Leuven friction model is modified and augmented in [8]. In the first modification described in [8], the state derivative is given by

$$\dot{z}(t) = \left(1 - \text{sgn}\left(\frac{f_h(z)}{f_{ss}(v)}\right) \left|\frac{f_h(z)}{f_{ss}(v)}\right|^n\right) v(t). \qquad (14)$$

Both the Leuven model (11) and the modified Leuven model (14) produce minor hysteresis loops that close in a nonlocal-memory-like way when the friction force $f_f(t)$ is plotted versus $z(t)$, as seen in Figure 13(a). However, the closure of the $f_f(t)$ versus $z(t)$ loops does not guarantee that hysteresis loops are closed when friction is plotted versus the externally measured position of the body $x(t)$, as seen in Figure 13(b). Decomposition of the total motion into elastic and plastic components in Figure 13(a) illuminates why the Leuven model shows closed-friction force-position curves when $f_f(t)$ is plotted versus $z(t)$ but not when $f_f(t)$ is plotted versus $x(t)$ in Figure 13(b). From (11), the rate of plastic displacement is given by

$$\dot{w}(t) = -\text{sgn}\left(\frac{f_d(z)}{f_{ss}(v) - f_b(t)}\right) \left|\frac{f_d(z)}{f_{ss}(v) - f_b(t)}\right|^n v(t), \quad (15)$$

and the total plastic displacement during a cycle of oscillatory motion is given by
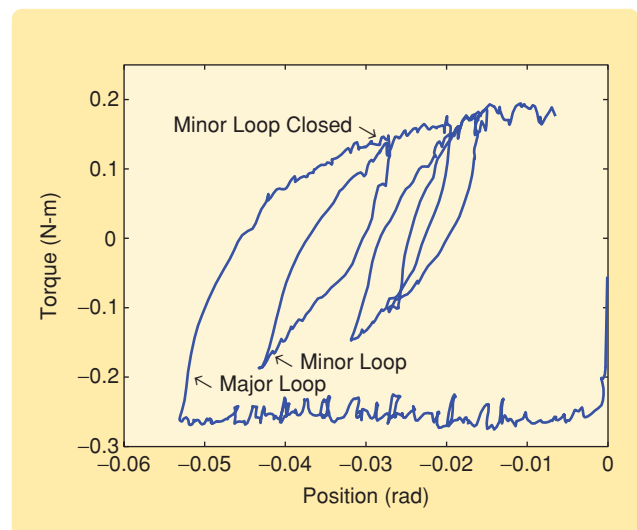

FIGURE 12 Experimental friction data obtained for a precision motion stage of an electrical discharge machine [7], [26]. The major hysteresis loop is the outer loop. Several minor hysteresis loops are also shown. If the trajectory returns to the major loop when a minor loop is closed, the hysteresis has nonlocal memory.

$$\Delta w = \int_{t_1}^{t_2} \dot{w}(t)\, dt + \int_{t_2}^{t_3} \dot{w}(t)\, dt, \qquad (16)$$

where t_1 and t_3 are the start and finish times of a cycle, and t_2 is the time of the intermediate reversal; see Figure 13. Equation (15) depends in a complicated way on the profiles of $f_{ss}(v(t))$ and $f_d(z(t))$ as well as the time histories of $v(t)$ and $z(t)$. In general, the plastic displacement over the first half-cycle does not exactly cancel the plastic displacement over the second half-cycle.

Figures 13 and 14 illustrate simulations of the sliding mass shown in Figure 5 with (11)–(13) using the parameters given in Table 2 and with Stribeck friction and a nonlinear spring $f_d(z)$ given as

$$f_{ss}(v) = \left[(f_S - f_C) \frac{1}{1 + (v(t)/v_s)^2} + f_C \right] \mathrm{sgn}\,(v(t)), \quad (17)$$

$$f_d(z) = \sin(\sigma_0 z). \qquad (18)$$

The force applied in the simulation is shown in Figure 14(a). Similar results are obtained using $f_d(z) = \sigma_0 z$ as well as with the modified Leuven model, using (14).

Considering the drift seen in Figure 14(b), the Z-properties chart for the Leuven model is presented in Figure 15. The structural change marked by the double heavy line at $z(t) = z_{ss}$ corresponds to resetting the hysteresis stacks upon entering steady sliding.

THE MAXWELL-SLIP MODEL

The MS model [21], [22], [29] is based on several elementary sliding contacts, called MS elements, operating in parallel within a friction junction. The elementary sliding contacts are described by

$$\dot{z}_i(t) = \begin{cases} v(t), & \text{sticking phase,} \\ 0, & \text{sliding phase,} \end{cases} \qquad (19)$$

where $z_i(t)$ is the state of the ith elementary sliding contact. A transition from sticking to sliding occurs when
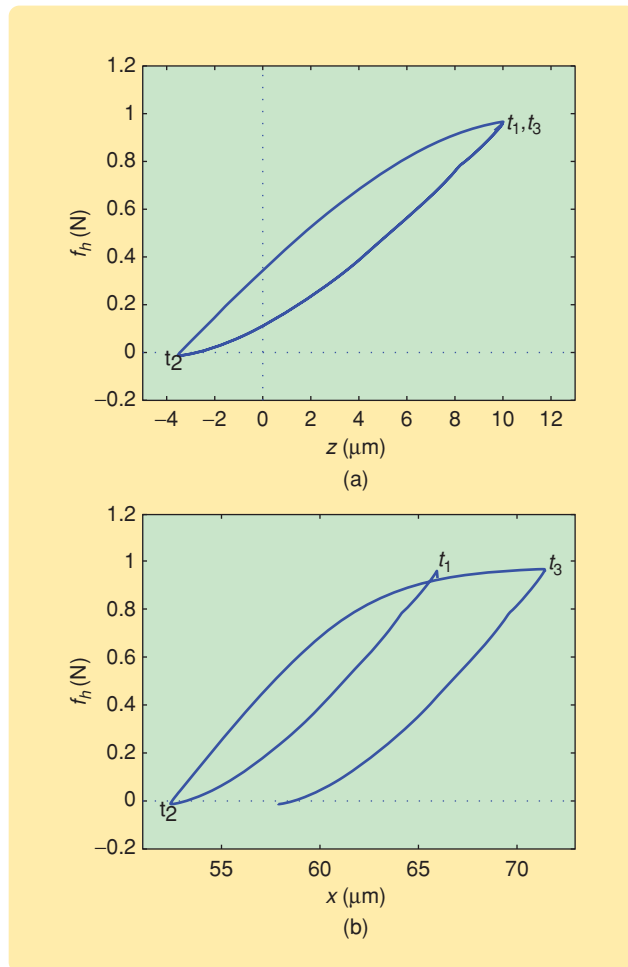

FIGURE 13 Friction simulated with the Leuven model. (a) shows f_h versus z, for which the model forms closed loops, while (b) shows f_h versus x, for which the loops are not closed. The open loops in (b) illustrate the absence of hysteresis with nonlocal memory, which requires that friction-position loops close, as seen in Figure 12.
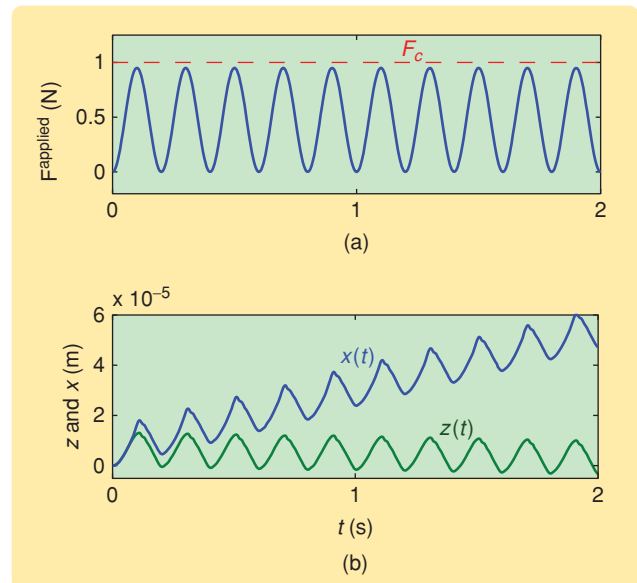
FIGURE 14 Applied force and internal state for the Leuven model simulation. The (a) simulated applied force, which comprises a bias and oscillatory component, never exceeds the Coulomb friction level. The externally measurable response $x(t)$ exhibits a steady net motion, that is, drift. Experiments with an electric servo show no net motion under similar conditions [4].

TABLE 2 Parameters used in the friction simulations shown in figures 13 and 14. The microdamping parameter σ_1 corresponds to a damping factor of 0.2 in the microdynamics.

Parameter	Value	Units	Parameter	Value	Units
M	1.0	kg	σ_0	10^5	N/m
f_C	1.0	N	σ_1	63	N-s/m
f_S	1.1	N	σ_2	0	N-s/m
n	7.0		v_s	10^{-5}	m/s

$|z_i(t)| = z_{ss}^i$, while a transition from sliding to sticking occurs upon a reversal of $v(t)$. Associated with each elementary sliding element is a stiffness σ_0^i and a friction level f_{ss}^i, with

$$z_{ss}^i = f_{ss}^i / \sigma_0^i, \qquad (20)$$

where the composite friction level in steady sliding is the sum of the elementary friction levels given by

$$f_{ss} = \sum_{i=1}^{N} f_{ss}^i .$$

The total friction force is given by

$$f_f(t) = \sum_{i=1}^{N} \sigma_0^i z_i(t) + f(v(t)), \qquad (21)$$

where N is the number of elementary contacts. The value $N = 10$ is used in the examples of [20]. Viscous friction is represented by a general function $f(v(t))$ [9]. The expression $f(v(t)) = \sigma_2 v(t)$ gives the special case of linear viscous friction.

When the elementary sliding contacts are combined, the composite stiffness for small deflections is the sum of the elementary stiffnesses, that is,

$$\sigma_0 = \sum_{i=1}^{N} \sigma_0^i .$$

The composite stiffness diminishes as the elementary contacts transition from the sticking phase to the sliding phase. Taking z_{ss} as the equivalent maximum asperity deflection in the Dahl or LuGre models, the terms z_{ss}^i are selected so that the maximum z_{ss}^i is equal to z_{ss}, and the minimum z_{ss}^i is much less than z_{ss}. In this way the elementary contact with the smallest value of z_{ss}^i transitions to sliding first, while complete transition to sliding is achieved when the contact having the maximum value of z_{ss}^i is sliding.

The Z-properties chart for the MS model is shown in Figure 16. For motions with $|z_1(t)| < z_A$, the model simulates an elastic contact with $\sigma_0 = \sum \sigma_0^i$, where $z_A = \min_i z_{ss}^i$. Since the MS model is purely elastic in this regime, the motion shows neither damping in the microdynamics nor hysteresis. Once the first element begins to slide, the model shows rate-independent hysteresis with nonlocal memory. When a velocity reversal occurs, the nonlocal memory effect is captured in the values $w_i(t)$ of the sliding elements, where $w_i(t)$ is given by $x(t) = z_i(t) + w_i(t)$. For $|z(t)| < z_{ss}$, at least one element is not sliding, preventing drift. Finally, when all elements are sliding, $\dot{z}_i(t) = 0$ according to the second case of (19). A double heavy line is used at $|z(t)| = z_{ss}$ because of the change of model structure corresponding to the transition
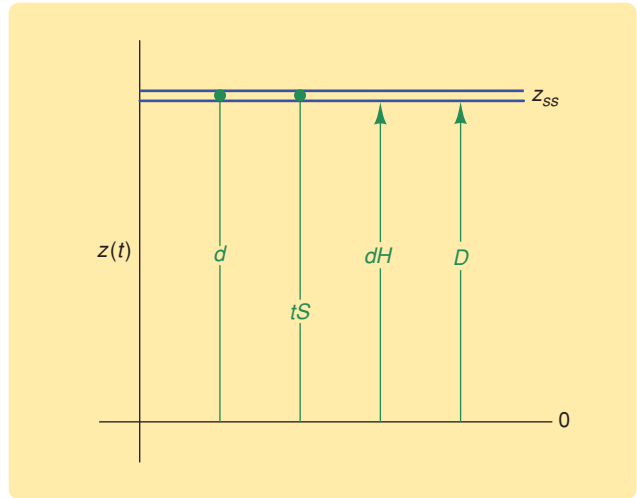
FIGURE 15 Z-properties chart for the Leuven and modified Leuven friction models. For $z(t) < z_{ss}$, the models exhibit drift (D) and rate-dependent hysteresis (dH) but not hysteresis with nonlocal memory. The change of structure illustrated by the double line corresponds to resetting the hysteresis stacks upon entering steady sliding.

FIGURE 16 Z-Properties chart for the Maxwell-slip model. The level z_A is the minimum of the z_{ss}^i of (20), while the level z_{ss} is the maximum of the z_{ss}^i. The model has an unusual regime for $z(t) < z_A$, where the microdynamic model reduces to a spring with stiffness σ_0. Neither hysteresis nor damping of the microdynamics is present in this regime.

> **A shared feature of these state-variable models is that each has an internal state or states with units of length or angle.**

from case 1 to case 2 of (19). It appears possible with the MS model to set the minimal z_{ss}^i to zero. In this case, $z_A = 0$, and the purely elastic interval is removed from the Z-properties chart.

THE GENERALIZED GMS MODEL

Because a variety of extensions to the basic MS model are called GMS models, the name G-GMS model is used in [9] for the model of (22)–(25). The G-GMS model simulates several elementary sliding contacts. Ten sliding contacts are used in the examples of [9]. Each sliding contact of the G-GMS model is described by

$$\dot{z}_i(t) = v(t) \qquad \text{sticking phase,} \quad (22)$$

$$\dot{z}_i(t) = \text{sgn}(v)C_i \left(1 - \frac{z_i(t)}{z_{ss}^i(v(t))}\right), \quad \text{sliding phase.} \quad (23)$$

As with the MS model, $z_i(t)$ is the state of the ith elementary sliding contact. To model Stribeck friction, the model is augmented with velocity-dependent $z_{ss}^i(v(t))$ and (23). The parameters C_i in (23), which are called the acceleration parameters in [9], determine how rapidly the terms $z_i(t)$ track changes in $z_{ss}^i(v(t))$, which are given by
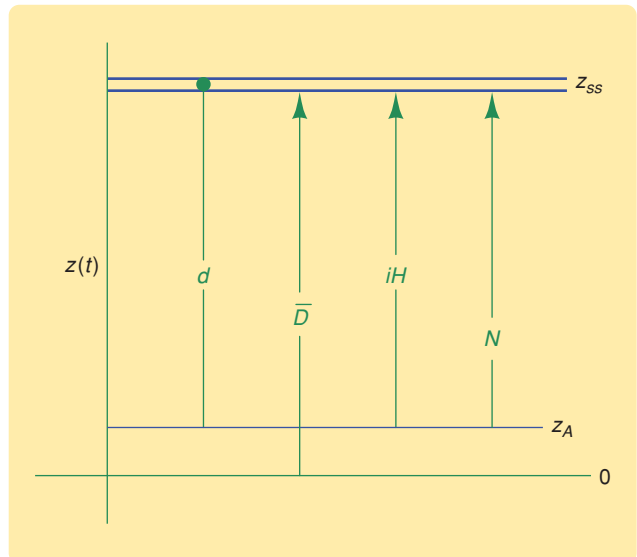

FIGURE 17 Z-Properties chart for the generalized generalized Maxwell-slip model. For $z(t) < z_{ss}$, the model exhibits damped microdynamics (d) and drift-free behavior ($\overline{D}$), while, for $z(t) > z_A$, the model exhibits rate-dependent hysteresis (dH) with nonlocal memory (N), where $z_A = \min_i z_{ss}^i$.

$$z_{ss}^i(v(t)) = f_{ss}^i(v(t))/\sigma_0^i. \quad (24)$$

The total friction force expression is also augmented in the G-GMS model to include microdamping, that is,

$$f_{\text{f}}(t) = \sum_{i=1}^{N} \left(\sigma_0^i z_i(t) + \sigma_1^i \dot{z}_i(t)\right) + f(v(t)). \quad (25)$$

The elementary sliding contacts combine by the same rules as for the MS model, with the addition that

$$f_{\text{ss}}(v(t)) = \sum_{i=1}^{N} f_{\text{ss}}^i(v(t)), \quad (26)$$

where the $f_{ss}^i(v(t))$ are chosen to reflect the Stribeck curve.

The Z-properties chart for the G-GMS model is shown in Figure 17. This chart is distinct from the Z-properties chart for the MS model in that the G-GMS model has damping in the microdynamics all the way down to $z(t) = 0$. Additionally, using the f_{ss}^i given in [9], the G-GMS model has rate-dependent hysteresis owing to the rate dependence in $z_{ss}^i(v(t))$. Like the MS model, the G-GMS model approximates hysteresis with nonlocal memory and does not show drift.

CONCLUSIONS

The principal dynamic friction models discussed in the control literature have the common feature of incorporating one or several state variables with units of displacement and representing elasticity. Beyond this similarity, however, the models show a wide variety of dynamic behavior. By showing the regimes of $z(t)$ on which various dynamic behaviors are manifested, the Z-properties chart is a graphical tool for comparing and contrasting model behavior. Carefully examining the dynamic behavior as a function of $z(t)$ is also a useful exercise for understanding the characteristics of each model. The examples presented here demonstrate the diversity of the dynamic behavior of the state-variable friction models and utility of the Z-properties chart.

ACKNOWLEDGMENTS

REFERENCES

[1] P. Dahl, "A solid friction model," Aerospace Corp., El Segundo, CA, Tech. Rep. TOR-0158(3107-18)-1, 1968.

[2] P. Dahl, "Solid friction damping of mechanical vibrations," *AIAA J.*, vol. 14, no. 12, pp. 1675–1682, 1976.

[3] D. Haessig and B. Friedland, "On the modeling and simulation of friction," *J. Dyn. Syst. Meas. Contr.*, vol. 113, no. 3, pp. 354–362, 1991.

[4] B. Armstrong-Hélouvry, *Control of Machines with Friction*. Norwell, MA: Kluwer, 1991.

[5] C. Canudas de Wit, H. Olsson, K. Aström, and P. Lischinsky, "A new model for control of systems with friction," *IEEE Trans. Automat. Contr.*, vol. 40, no. 3, pp. 419–425, 1995.

[6] J. Swevers, F. Al-Bender, C. Ganseman, and T. Projogo, "An integrated friction model structure with improved presliding behavior for accurate friction compensation," *IEEE Trans. Automat. Contr.*, vol. 45, no. 4, pp. 675–686, 2000.

[7] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter, "Single state elasto-plastic friction models," *IEEE Trans. Automat. Contr.*, vol. 47, no. 5, pp. 787–792, 2002.

[8] V. Lampaert, J. Swevers, and F. Al-Bender, "Modification of the Leuven integrated friction model structure," *IEEE Trans. Automat. Contr.*, vol. 47, no. 4, pp. 683–687, 2002.

[9] F. Al-Bender, V. Lampaert, and J. Swevers, "The generalized Maxwell-slip model: A novel model for friction simulation and compensation," *IEEE Trans. Automat. Contr.*, vol. 50, no. 11, pp. 1883–1887, 2005.

[10] A. Tustin, "The effects of backlash and of speed-dependent friction on the stability of closed-cycle control systems," *IEE J.*, vol. 94, part 2A, pp. 143–151, 1947.

[11] F. Al-Bender, V. Lampaert, and J. Swevers, "Modeling of dry sliding friction dynamics: From heuristic models to physically motivated models and back," *Chaos*, vol. 14, no. 2, pp. 446–460, 2004.

[12] B. Armstrong-Hélouvry, P. Dupont, and C. Canudas de Wit, "A survey of models, analysis tools and compensation methods for the control of machines with friction," *Automatica*, vol. 30, no. 7, pp. 1083–1138, 1994.

[13] H. Olsson, K.J. Astrom, C. Canudas De Wit, M. Gafvert, and P. Lischinsky, "Friction models and friction compensation," *Eur. J. Contr.*, vol. 4, no. 3, pp. 176–195, 1998.

[14] P. Bliman and M. Sorine, "Friction modelling by hysteresis operators. application to Dahl, stiction and Stribeck effects," in *Proc. Conf. Models of Hysteresis*, A. Visintin Ed. (Pitman Research Notes in Mathematics), Trento, Italy, 1991, pp. 10–19.

[15] V. Hayward and B. Armstrong, "A new computational model of friction applied to haptic rendering," in *Preprints of ISER'99, Experimental Robotics VI*, P. Corke and J. Trevelyan, Eds. New York: Springer-Verlag, 2000, pp. 403–412.

[16] F. Altpeter, M. Grunenberg, P. Myszkorowski, and R. Longchamp, "Auto-tuning of feedforward friction compensation based on the gradient method," in *Proc. 2000 American Control Conf.*, vol. 4. Chicago, IL, 2000, pp. 2600–2604.

[17] P. Dupont, "The effect of Coulomb friction on the existence and uniqueness of the forward dynamics problem," in *Proc. 1992 Int. Conf. Robotics and Automation*, Nice, France, 1992, pp. 1442–1447.

[18] P. Dupont, "The effect of friction on the forward dynamics problem," *Int. J. Robot. Res.*, vol. 12, no. 2, pp. 164–179, 1993.

[19] D. Threlfall, "The inclusion of coulomb friction in mechanisms programs with particular reference to dram," *Mechanism Mach. Theory*, vol. 13, no. 4, pp. 475–483, 1978.

[20] V. Lampaert, F. Al-Bender, and J. Swevers, "A generalized Maxwell-slip friction model appropriate for control purposes," in *Proc. IEEE Int. Conf. Physics and Control*, St. Petersburg, 2003, pp. 20–22.

[21] D.D. Rizos and S.D. Fassois, "Presliding friction identification based upon the Maxwell slip model structure," *Chaos*, vol. 14, no. 2, pp. 431–445, 2004.

[22] J.H. Oh, A.K. Padthe, D.S. Bernstein, D.D. Rizos, and S.D. Fassois, "Duhem models for hysteresis in sliding and presliding friction," in *Proc. 44th IEEE Conf. Decision and Control and 2005 European Control Conf.*, Dec. 2005, pp. 8132–8137.

[23] S. Futami, A. Furutani, and S. Yoshida, "Nanometer position and its micro-dynamics," *Nanotechnology*, vol. 1, no. 1, pp. 31–37, 1990.

[24] R. Stribeck, "Die wesentlichen eigenschaften der gleit- und rollen-lager—the key qualities of sliding and roller bearings," *Zeitschrift des Vereines Seutscher Ingenieure*, vol. 46, no. (38, 39), pp. 1342–1348, 1432–1437, 1902.

[25] B. Armstrong, "Friction: Experimental determination, modeling and compensation," in *Proc. 1988 Int. Conf. Robotics and Automation*, Philadelphia, PA, 1988, pp. 1422–1427.

[26] F. Altpeter, "Friction modeling, identification and compensation," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, Switzerland, 1999 [Online]. Available: http://biblion.epfl.ch/EPFL/theses/1999/1988/EPFL_TH1988.pdf.

[27] P. Berenyi, G. Horvath, V. Lampaert, and J. Swevers, "Nonlocal hysteresis function identification and compensation with neural networks," *IEEE Trans. Instrum. Meas.*, vol. 54, pp. 2227–2238, Dec. 2005.

[28] V. Lampaert and J. Swevers, "Online identification of hysteresis functions with non-local memory," in *Proc. 2001 IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics*, vol. 2. Como, Italy, 2001, pp. 833–837.

[29] U. Parlitz, A. Hornstein, D. Engster, F. Al-Bender, V. Lampaert, T. Tjahjowidodo, S.D. Fassois, D. Rizos, C.X. Wong, K. Worden, and G. Manson, "Identification of pre-sliding friction dynamics," *Chaos*, vol. 14, no. 2, pp. 420–430, 2004.

AUTHOR INFORMATION

Brian S.R. Armstrong (bsra@uwm.edu) received the B.Sc. in physics and mechanical engineering from the Massachusetts Institute of Technology, Cambridge, and the M.Sc. and Ph.D. in electrical engineering from Stanford University, Stanford, California. He is a Senior Member of the IEEE. Since 1995, he has been an associate professor in the Electrical Engineering and Computer Science Department of the University of Wisconsin-Milwaukee. He has worked closely with industry, including visiting positions with the Innovation Center of Eaton Corporation, Milwaukee (2005–2006), the Mitsubishi Electric Information Technology Center, Europe, Rennes, France (1997–1998), and the Mitsubishi Electric Corporation Product Development laboratory, Amagasaki, Japan (1986). He is the cofounder and past CEO of MetriCam, Inc., a technology startup company focusing on image metrology. He is the author of more than 50 books and book chapters, papers, and conference presentations in friction modeling, identification and compensation, robotics, force control, and spatial measurement from images and has several U.S. patents disclosing mechanisms, Doppler-radar-based 3-D tracking, and methods of image metrology. His most recent book is *Precision Landmark Location for Machine Vision and Photogrammetry* (New York: Springer, 2007). He is a member of Tau Beta Pi and received the UWM University Outstanding Teaching Award, the MIT/DeFlorez design award, and a Hewlett-Packard Faculty Development Program Fellowship. He is a registered professional engineer. He can be contacted at the Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, 3200 N Cramer St., Milwaukee, WI 53211 USA.

Qunyi Chen received the M.Sc. in mechanical engineering and in electrical engineering from North Carolina State University, Raleigh, in 1998 and 1999, respectively. He is currently pursuing a Ph.D. in mechanical engineering at North Carolina State University. His research interests include precision motion control and design of electromechanical system. He worked for GE Healthcare from 2000 to 2004, designing and developing electromechanical subsystems for medical diagnostic X-ray equipment. His research interest includes design and control of electromechanical system, and precision motion control.

Duhem Modeling of Friction-Induced Hysteresis

ASHWANI K. PADTHE, BOJANA DRINCIC, JINHYOUNG OH, DEMOSTHENIS D. RIZOS,
SPILIOS D. FASSOIS, and DENNIS S. BERNSTEIN

EXPERIMENTAL DETERMINATION OF GEARBOX STICTION

Friction is a dynamic phenomenon of widespread importance, and the associated literature is vast; overviews are given in [1]–[5]. Friction can be viewed as an emergent, macroscopic phenomenon arising from molecular interaction. Consequently, both physical (physics-based) and empirical (experiment-based) models have been studied [2], [6]–[14]. Estimation and control methods are available for applications involving friction [15]–[18]; however, these topics are beyond the scope of this article.

Friction models distinguish between *presliding friction* and *sliding friction*. Presliding or *micro-slip* friction refers to the friction forces that occur when the relative displacement between two contacting surfaces is microscopic, that is, on the order of the asperities (roughness features) on the surfaces. Sliding friction refers to the friction forces that arise when the relative displacement is macroscopic. Understanding presliding friction is useful for high precision motion control applications. For example, hysteresis can occur between the presliding friction force input and the displacement output [7], [11], [12].

From a mathematical point of view, friction modeling is challenging since these models often involve nonsmooth dynamics. For example, the most widely used dry friction model, namely, Coulomb friction, is discontinuous. Additional discontinuous dry friction models are studied in [19]. Some friction models are continuous but have non-Lipschitzian dynamics, which is a necessary condition for

finite-settling-time behavior and the associated lack of time-reversibility [20], [21]. Table 1 classifies the properties of some widely used friction models.

Hysteresis is the result of *multistability*, which refers to the existence of multiple attracting equilibria [22]–[24]. Multistability implies that hysteresis is a quasi-static phenomenon in the sense that the hysteresis map is the limit of a sequence of periodic dynamic input-output maps as the period of the input increases without bound. In both presliding and sliding friction models, there exist multiple equilibria corresponding to states that correspond to constant friction forces under constant displacement or velocity.

In this article we examine several classical friction models from a hysteresis modeling point of view and study the
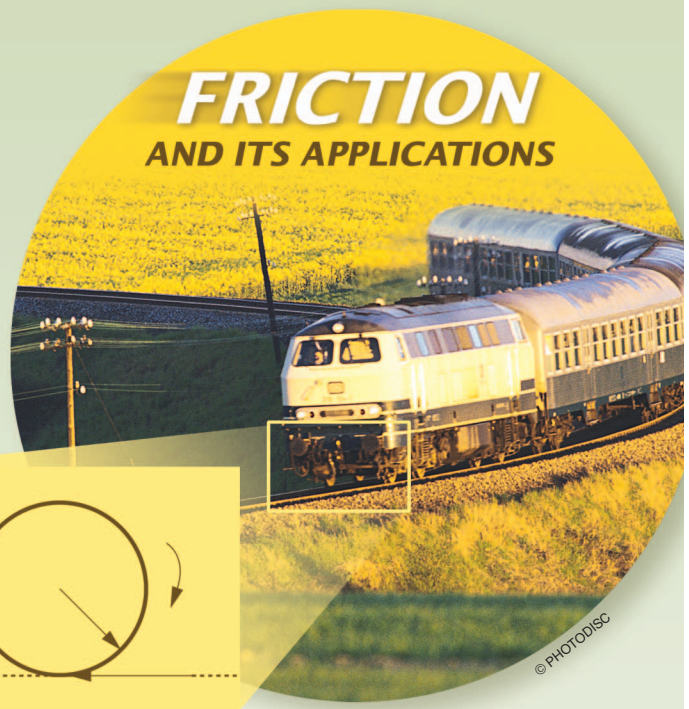
hysteresis induced by these friction models when incorporated into a physical system. Our starting point is [25], which focuses on the Duhem model for hysteresis. For more information on Duhem models, see [26]. The Duhem model has the property that, under constant inputs, every state is an equilibrium. When there exist multiple attracting (step-convergent) equilibria for a given step input, the system exhibits hysteresis under inputs that drive the system through distinct equilibria that map into distinct outputs. In certain cases, the limiting input-output map is independent of the input period; this case is known as *rate-independent* hysteresis. In general, the hysteresis map is *rate dependent*, although the terminology is slightly misleading since, as already noted, hysteresis per se is a quasi-static phenomenon.

The generalized Duhem model $\dot{x} = f(x, u)g(\dot{u})$ and its specialization $\dot{x} = (Ax + Bu)g(\dot{u})$, known as the *semilinear Duhem model*, are considered in [25]. These models give rise to rate-independent hysteresis when the function g is positively homogeneous; otherwise, the hysteresis is generally rate dependent.

In the present article we consider three friction models, namely, the Dahl, LuGre (Lund/Grenoble), and Maxwell-slip models. We recast each model in the form of a generalized or semilinear Duhem model to provide a unified framework for comparing the hysteretic nature of these models. For example, the Dahl model is shown to be a rate-independent generalized Duhem model. Furthermore, in one special case, the Dahl model is also a semilinear Duhem model for which a closed-form solution is available. Similarly, the LuGre model is a rate-dependent generalized Duhem model. Next, we embed each friction model within a single-degree-of-freedom mechanical model to examine and compare the hysteretic response of the combined system.

Finally, we develop an experimental testbed for friction identification. The testbed consists of a dc motor with a speed-reduction gearhead, encoder measurements of the shaft, tachometer measurements of the shaft angular velocity, and load-cell tension measurements of a cable wound around the drum. By operating this testbed under quasi-static conditions, we compare its hysteretic response to the simulated response of the system under various friction models. The goal is to identify a model for the friction and stiction effects observed in the testbed by comparing the simulation and experimental results.

The objective of this article is to reformulate the Dahl, LuGre, and Maxwell-slip models as Duhem models to understand their hysteretic properties. This classification provides the framework for identifying a friction model that captures the hysteretic behavior of the motor gearbox.

The contents of the article are as follows. In the following section we review the basic theory of the Duhem model. Next, we recast the Dahl, LuGre, and Maxwell-slip models as Duhem models and relate their dynamic behavior to properties of the Duhem models. We then study the sliding friction dynamics of the three friction models. Next, we consider friction-induced hysteresis in a mass-spring system. This system is studied as a special case of a linear time-invariant system with Duhem feedback. We then develop a model of the experimental setup and simulate the model using all three friction models. We then report the experimental results and compare them with the simulation results to obtain estimates of the friction parameters. Finally, we give some concluding remarks.

GENERALIZED AND SEMILINEAR DUHEM MODELS

In this section, we summarize the main results of [25] concerning the generalized and semilinear Duhem models. Consider the single-input, single-output *generalized Duhem model*

$$\dot{x}(t) = f(x(t), u(t))g(\dot{u}(t)), \quad x(0) = x_0, \quad t \geq 0, \quad (1)$$

$$y(t) = h(x(t), u(t)), \quad (2)$$

where $x : [0, \infty) \rightarrow \mathbb{R}^n$ is absolutely continuous, $u : [0, \infty) \rightarrow \mathbb{R}$ is continuous and piecewise C^1, $f : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^{n \times r}$ is continuous, $g : \mathbb{R} \rightarrow \mathbb{R}^r$ is continuous and satisfies $g(0) = 0$, and $y : [0, \infty) \rightarrow \mathbb{R}$, and $h : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ are continuous. The value of $\dot{x}(t)$ at a point t at which $\dot{u}(t)$ does not exist can be assigned arbitrarily. We assume that the solution to (1) exists and is unique on all finite intervals. Under these assumptions, x and y are continuous and piecewise C^1. The terms closed

TABLE 1 Classification and properties of friction models. Each friction model is a Duhem model, with either rate-independent or rate-dependent dynamics. Non-Lipschitzian dynamics is a necessary condition for finite-settling-time convergence.

Friction Model		Duhem Type	Rate Dependence	Continuity
Coulomb		Static	Rate independent	Discontinuous
Dahl	$\gamma = 0$	Generalized	Rate independent	Discontinuous
	$0 < \gamma < 1$	Generalized	Rate independent	Continuous but not Lipschitz
	$\gamma = 1$	Semilinear	Rate independent	Lipschitz
	$\gamma > 1$	Generalized	Rate independent	Lipschitz
LuGre		Generalized	Rate dependent	Lipschitz
Maxwell-slip		Generalized	Rate independent	Discontinuous

curve, limiting periodic input-output map, hysteresis map, and rate independence are defined as follows.

Definition 1
The nonempty set $\mathcal{H} \subset \mathbb{R}^2$ is a closed curve if there exists a continuous, piecewise C^1, and periodic map $\gamma : [0, \infty) \to \mathbb{R}^2$ such that $\gamma([0, \infty)) = \mathcal{H}$.

Definition 2
Let $u : [0, \infty) \to [u_{\min}, u_{\max}]$ be continuous, piecewise C^1, periodic with period α, and have exactly one local maximum $u_{\max}$ in $[0, \alpha)$ and exactly one local minimum $u_{\min}$ in $[0, \alpha)$. For all $T > 0$, define $u_T(t) \triangleq u(\alpha t/T)$, assume that there exists $x_T : [0, \infty) \to \mathbb{R}^n$ that is periodic with period T and satisfies (1) with $u = u_T$, and let $y_T : [0, \infty) \to \mathbb{R}$ be given by (2) with $x = x_T$ and $u = u_T$. For all $T > 0$, the periodic input-output map $\mathcal{H}_T(u_T, y_T, x_0)$ is the closed curve $\mathcal{H}_T(u_T, y_T, x_0) \triangleq \{(u_T(t), y_T(t)) : t \in [0, \infty)\}$, and the limiting periodic input-output map $\mathcal{H}_\infty(u, x_0)$ is the closed curve $\mathcal{H}_\infty(u, x_0) \triangleq \lim_{T \to \infty} \mathcal{H}_T(u_T, y_T, x_0)$ if the limit exists. If there exist $(u, y_1), (u, y_2) \in \mathcal{H}_\infty(u, x_0)$ such that $y_1 \neq y_2$, then $\mathcal{H}_\infty(u)$ is a hysteresis map, and the generalized Duhem model is hysteretic. If, in addition $\mathcal{H}_\infty(u, x_0)$ is independent of x_0 then the generalized Duhem model has local memory, and we write $\mathcal{H}_\infty(u)$. Otherwise, $\mathcal{H}_\infty(u, x_0)$ has nonlocal memory.

Definition 3
The continuous and piecewise C^1 function $\tau : [0, \infty) \to [0, \infty)$ is a positive time scale if $\tau(0) = 0$, τ is nondecreasing, and $\lim_{t \to \infty} \tau(t) = \infty$. The generalized Duhem model (1), (2) is rate independent if, for every pair of continuous and piecewise C^1 functions x and u satisfying (1) and for every positive time scale τ, it follows that $x_\tau(t) \triangleq x(\tau(t))$ and $u_\tau(t) \triangleq u(\tau(t))$ also satisfy (1).

The following result is proved in [25].

Proposition 1
Assume that g is positively homogeneous, that is, $g(\alpha v) = \alpha g(v)$ for all $\alpha > 0$ and $v \in \mathbb{R}$. Then the generalized Duhem model (1), (2) is rate independent.

If g is positively homogeneous, then there exist $h_+, h_- \in \mathbb{R}^r$ such that

$$g(v) = \begin{cases} h_+ v, & v \geq 0, \\ h_- v, & v < 0, \end{cases} \tag{3}$$

and the rate-independent generalized Duhem model (1), (2) can be reparameterized in terms of u [25]. Specifically, consider

$$\frac{\mathrm{d}\hat{x}(u)}{\mathrm{d}u} = \begin{cases} f_+(\hat{x}(u), u), & \text{when } u \text{ increases}, \\ f_-(\hat{x}(u), u), & \text{when } u \text{ decreases}, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

$$\hat{y}(u) = h(\hat{x}(u), u), \tag{5}$$

for $u \in [u_{\min}, u_{\max}]$ and with initial condition $\hat{x}(u_0) = x_0$, where $f_+(x, u) \triangleq f(x, u)h_+$, $f_-(x, u) \triangleq f(x, u)h_-$, and $u_0 \in [u_{\min}, u_{\max}]$. Then $x(t) \triangleq \hat{x}(u(t))$ and $y(t) \triangleq \hat{y}(u(t))$ satisfy (1), (2). Note that the reparameterized Duhem model (4) and (5) can be viewed as a time-varying dynamical system with nonmonotonic time u.

As a specialization of (1) and (2), we now consider the *rate-independent semilinear Duhem model*

$$\dot{x}(t) = [\, \dot{u}_+(t)I_n \quad \dot{u}_-(t)I_n \,]$$
$$\times \left(\begin{bmatrix} A_+ \\ A_- \end{bmatrix} x(t) + \begin{bmatrix} B_+ \\ B_- \end{bmatrix} u(t) + \begin{bmatrix} E_+ \\ E_- \end{bmatrix} \right), \tag{6}$$

$$y(t) = Cx(t) + Du(t), \qquad x(0) = x_0, \quad t \geq 0, \tag{7}$$

where $A_+ \in \mathbb{R}^{n \times n}$, $A_- \in \mathbb{R}^{n \times n}$, $B_+ \in \mathbb{R}^n$, $B_- \in \mathbb{R}^n$, $E_+ \in \mathbb{R}^n, E_- \in \mathbb{R}^n, C \in \mathbb{R}^{1 \times n}, D \in \mathbb{R}$, and

$$\dot{u}_+(t) \triangleq \max\{0, \dot{u}(t)\}, \quad \dot{u}_-(t) \triangleq \min\{0, \dot{u}(t)\}. \tag{8}$$

Let $\rho(A)$ denote the spectral radius of $A \in \mathbb{R}^{n \times n}$ and let the *limiting input-output map* $\mathcal{F}_\infty(u, y)$ be the set of points $z \in \mathbb{R}^2$ such that there exists an increasing, divergent sequence $\{t_i\}_{i=1}^\infty$ in $[0, \infty)$ satisfying

$$\lim_{i \to \infty} \|(u(t_i), y(t_i)) - z\| = 0.$$

The following result given in [25] provides a sufficient condition for the existence of the limiting periodic input-output map for the rate-independent semilinear Duhem model.

Theorem 1
Consider the rate-independent semilinear Duhem model (6), (7), where $u : [0, \infty) \to [u_{\min}, u_{\max}]$ is continuous, piecewise C^1, and periodic with period α and has exactly one local maximum $u_{\max}$ in $[0, \alpha)$ and exactly one local minimum $u_{\min}$ in $[0, \alpha)$. Furthermore, define $\beta \triangleq u_{\max} - u_{\min}$, and assume that

$$\rho(e^{\beta A_+} e^{-\beta A_-}) < 1. \tag{9}$$

Then, for all $x_0 \in \mathbb{R}^n$, (6) has a unique periodic solution $x : [0, \infty) \to \mathbb{R}^n$, and the limiting periodic input-output map $\mathcal{H}_\infty(u, x_0)$ exists. Specifically, if A_+ and A_- are invertible, then

$$\mathcal{H}_\infty(u) = \{(u, \hat{y}_+(u)) \in \mathbb{R}^2 : u \in [u_{\min}, u_{\max}]\}$$
$$\cup \{(u, \hat{y}_-(u)) \in \mathbb{R}^2 : u \in [u_{\min}, u_{\max}]\}, \tag{10}$$

where

$$\hat{y}_+(u) = Ce^{A_+(u-u_{\min})}\hat{x}_+ - C\mathcal{Z}_+(u, u_{\min}) + Du,$$
$$\hat{y}_-(u) = Ce^{A_-(u-u_{\max})}\hat{x}_- - C\mathcal{Z}_-(u, u_{\max}) + Du,$$

and

$$\hat{x}_+ \triangleq -(I - e^{-\beta A_-} e^{\beta A_+})^{-1}(e^{-\beta A_-} \mathcal{Z}_+(u_{\max}, u_{\min})$$
$$+ \mathcal{Z}_-(u_{\min}, u_{\max})),$$
$$\hat{x}_- \triangleq -(I - e^{\beta A_+} e^{-\beta A_-})^{-1}(e^{\beta A_+} \mathcal{Z}_-(u_{\min}, u_{\max})$$
$$+ \mathcal{Z}_+(u_{\max}, u_{\min})),$$
$$\mathcal{Z}_+(u, u_0) \triangleq A_+^{-1}(uI - u_0 e^{A_+(u-u_0)})B_+$$
$$+ A_+^{-2}(I - e^{A_+(u-u_0)})B_+ + A_+^{-1}(I - e^{A_+(u-u_0)})E_+,$$
$$\mathcal{Z}_-(u, u_0) \triangleq A_-^{-1}(uI - u_0 e^{A_-(u-u_0)})B_-$$
$$+ A_-^{-2}(I - e^{A_-(u-u_0)})B_- + A_-^{-1}(I - e^{A_-(u-u_0)})E_-.$$

See [25] for the case in which A_+ or A_- is singular.

Definition 2 and Theorem 1 imply that the rate-independent semilinear Duhem model has local memory since the hysteresis map $\mathcal{H}_\infty(u)$ given by (10) is independent of the initial condition $x_0 \in \mathbb{R}^n$.

FRICTION MODELS

Dahl Model

The Dahl model [6], [27], [28] has the form

$$\dot{F}(t) = \sigma \left| 1 - \frac{F(t)}{F_C} \mathrm{sgn}\, \dot{u}(t) \right|^\gamma \mathrm{sgn}\left(1 - \frac{F(t)}{F_C} \mathrm{sgn}\, \dot{u}(t) \right) \dot{u}(t), \tag{11}$$

where F is the friction force, u is the relative displacement between the two surfaces in contact, $F_C > 0$ is the Coulomb friction force, $\gamma \geq 0$ is a parameter that determines the shape of the force-deflection curve (as represented by a plot of the friction force versus the relative displacement), and $\sigma > 0$ is the rest stiffness, that is, the slope of the force-deflection curve when $F = 0$. The right-hand side of (11) is Lipschitz continuous in F for $\gamma \geq 1$ but not Lipschitz continuous in F for $0 \leq \gamma < 1$.

When u is increasing, $\dot{F}(t)$ given by (11) is positive for all $F(t) < F_C$ and negative for all $F(t) > F_C$. Similarly, when u is decreasing, $\dot{F}(t)$ given by (11) is positive for all $F(t) < -F_C$ and negative for all $F(t) > -F_C$. Hence the magnitude of the friction force $F(t)$ approaches F_C under monotonic inputs. As shown in Figure 1, the parameter γ determines the shape of the hysteresis map. In practice, γ is typically set to zero or one. As shown in Figure 2, the friction force given by the Dahl model lags the friction force given by the Coulomb model when the direction of motion is reversed.

To represent (11) as a Duhem model, let

$$\mathcal{D}_+(F) \triangleq \sigma \left| 1 - \frac{F}{F_C} \right|^\gamma \mathrm{sgn}\left(1 - \frac{F}{F_C} \right), \tag{12}$$

$$\mathcal{D}_-(F) \triangleq \sigma \left| 1 + \frac{F}{F_C} \right|^\gamma \mathrm{sgn}\left(1 + \frac{F}{F_C} \right). \tag{13}$$
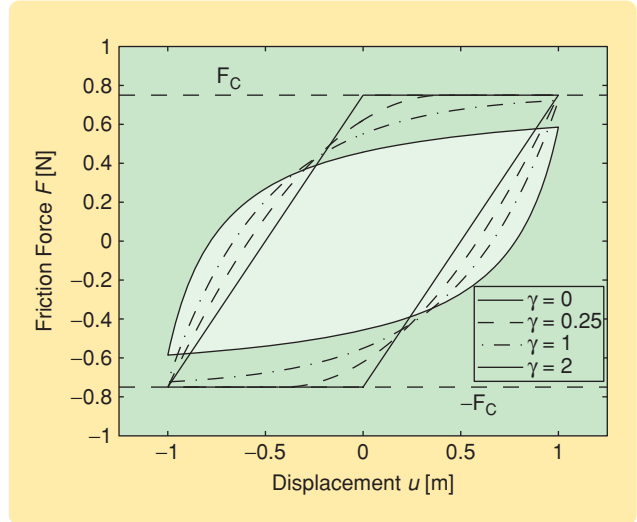
Then the Dahl model (11) can be rewritten as

FIGURE 1 Displacement u versus friction force F for hysteresis maps of the Dahl model for several values of γ. The shape of the hysteresis map from u to F depends on the value of γ. The numerical values used are $F_C = 0.75$ N, $\sigma = 1.5$ N/m, and $u(t) = \sin 0.1t$ m.
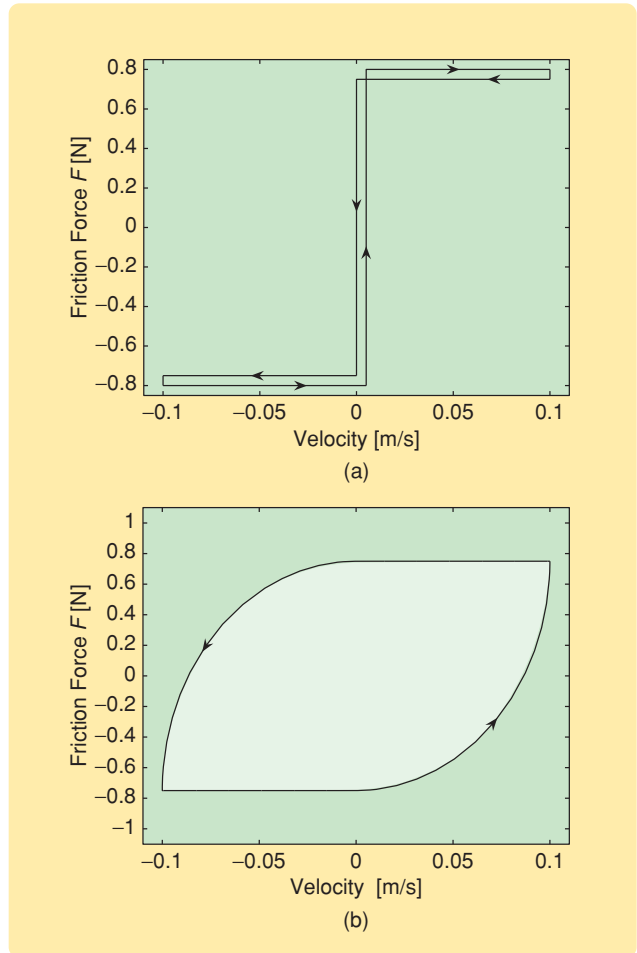

FIGURE 2 Velocity $\dot{u}$ versus friction force F for (a) the Coulomb model (with the traversals offset for clarity) and (b) the Dahl model with $\gamma = 0$. The friction force of the Dahl model lags the friction force of the Coulomb model when $\dot{u}$ changes sign.

$$\dot{F}(t) = \sigma \left[\, \mathcal{D}_+(F(t)) \quad \mathcal{D}_-(F(t)) \, \right]$$
$$\times \begin{bmatrix} \dot{u}_+(t) \\ \dot{u}_-(t) \end{bmatrix}, \tag{14}$$
$$y = F, \tag{15}$$

which, for all $\gamma \geq 0$, is a generalized Duhem model of the form (1), (2). Furthermore, since $g(\dot{u}) = [\, \dot{u}_+(t) \quad \dot{u}_-(t) \,]^T$ is positively homogeneous, Proposition 1 implies that (14) is rate independent for all $\gamma \geq 0$.

Let $\gamma = 1$. Then (11) becomes

$$\dot{F}(t) = \sigma \left(1 - \frac{F(t)}{F_C} \operatorname{sgn} \dot{u}(t) \right) \dot{u}(t)$$
$$= \left[-\frac{\sigma}{F_C} F(t) + \sigma \quad \frac{\sigma}{F_C} F(t) + \sigma \right] \begin{bmatrix} \dot{u}_+(t) \\ \dot{u}_-(t) \end{bmatrix},$$

which is a rate-independent semilinear Duhem model. Furthermore, the convergence condition (9) becomes

$$e^{-2\frac{\beta\sigma}{F_C}} < 1, \tag{16}$$

which holds if and only if $\beta > 0$. As a direct consequence of Theorem 1, which explicitly characterizes the hysteresis map, we have the following result. The corresponding hysteresis map is shown in Figure 1.

Corollary 1

Consider the Dahl model (11) with $\gamma = 1$. Let $u : [0, \infty) \to [u_{\min}, u_{\max}]$ be continuous, piecewise C^1, and periodic with period α and have exactly one local maximum $u_{\max}$ in $[0, \alpha)$ and exactly one local minimum $u_{\min}$ in
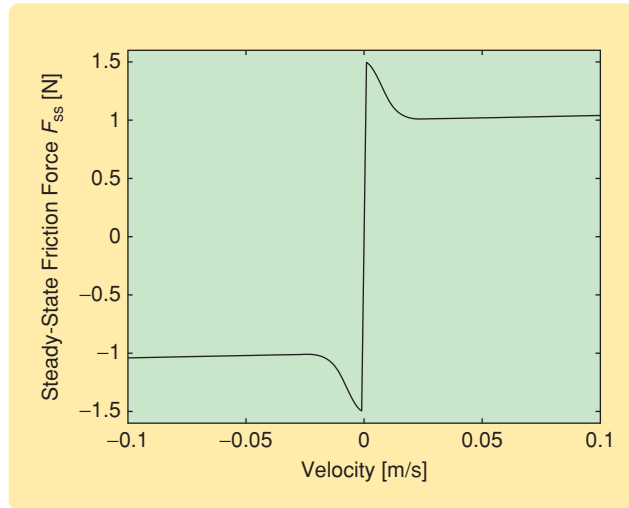

FIGURE 3 Steady-state friction force (21) given by the LuGre model. The drop in the friction force at low velocities is the Stribeck effect, while the Stribeck velocity $v_S = 0.001$ m/s is the velocity at which the steady-state friction force begins to decrease when the velocity is positive and increasing. The numerical values are $F_C = 1$ N, $F_S = 1.5$ N, $v_S = 0.001$ m/s, $\sigma_0 = 10^5$ N/m, $\sigma_1 = \sqrt{10^5}$ N-s/m, and $\sigma_2 = 0.4$ N-s/m.

$[0, \alpha)$. Then (16) holds, and (14), (15) has a unique periodic solution $F : [0, \infty) \to \mathbb{R}^n$, and, for all $x_0 \in \mathbb{R}^n$, the limiting periodic input-output map $\mathcal{H}_\infty(u)$ exists. Furthermore,

$$\mathcal{H}_\infty(u) = \left\{ (u, \hat{F}_+(u)) \in \mathbb{R}^2 : u \in [u_{\min}, u_{\max}] \right\}$$
$$\cup \left\{ (u, \hat{F}_-(u)) \in \mathbb{R}^2 : u \in [u_{\min}, u_{\max}] \right\}, \tag{17}$$

where

$$\hat{F}_+(u) \triangleq e^{-\frac{\sigma}{F_C}(u - u_{\min})} \hat{\alpha}_+ + F_C \left(1 - e^{-\frac{\sigma}{F_C}(u - u_{\min})} \right),$$
$$\hat{F}_-(u) \triangleq e^{\frac{\sigma}{F_C}(u - u_{\max})} \hat{\alpha}_- - F_C \left(1 - e^{-\frac{\sigma}{F_C}(u - u_{\max})} \right),$$

and

$$\hat{\alpha}_+ = -\hat{\alpha}_- = F_C \frac{e^{\frac{-\beta\sigma}{F_C}} - 1}{e^{-\frac{\beta\sigma}{F_C}} + 1}.$$

Corollary 1 implies that the Dahl model (11) with $\gamma = 1$ has local memory since it is a rate-independent semilinear Duhem model and $\mathcal{H}_\infty(u)$ defined by (17) is independent of x_0.

LuGre Model

The LuGre model [10], which models the asperities of two surfaces as elastic bristles, is given by

$$\dot{x}(t) = \dot{u}(t) - \frac{|\dot{u}(t)|}{r(\dot{u}(t))} x(t), \tag{18}$$
$$F(t) = \sigma_0 x(t) + \sigma_1 \dot{x}(t) + \sigma_2 \dot{u}(t), \tag{19}$$

where x is the average deflection of the bristles, u is the relative displacement, F is the friction force, and $\sigma_0, \sigma_1, \sigma_2 > 0$ are stiffness, damping, and viscous friction coefficients, respectively. The right-hand side of (18) is Lipschitz continuous with respect to x. In [1] and [10], $r(\dot{u}(t))$ is defined by

$$r(\dot{u}(t)) = \frac{F_C}{\sigma_0} + \frac{F_S - F_C}{\sigma_0} e^{-(\dot{u}(t)/v_S)^2}, \tag{20}$$

where $F_C > 0$ is the Coulomb friction force, F_S is the stiction (sticking friction) force, and v_S is the Stribeck velocity.

For a given constant velocity $\dot{u}$, the steady-state friction force F_{ss} obtained from (18) and (19) is

$$F_{ss}(\dot{u}) = \sigma_0 r(\dot{u}) \operatorname{sgn}(\dot{u}) + \sigma_2 \dot{u}. \tag{21}$$

The drop in friction force (see Figure 3) at low magnitudes of velocity is due to the Stribeck effect, while the Stribeck velocity is the velocity at which the steady-state friction force begins to decrease when the velocity is positive and increasing. The LuGre model (18), (19) combines the friction lag of the Dahl model with the Stribeck effect [10], [29].

Letting $F_S = F_C$ in (20) and $\sigma_1 = \sigma_2 = 0$ in 19, the LuGre model (18), (19), (20) is equivalent to the Dahl model (11)

with $\gamma = 1$ and $\sigma = 1$. With $y = F$, the state equations (18) and (19) can be written as

$$\dot{x}(t) = [\,1 \quad x(t)\,] \begin{bmatrix} \dot{u}(t) \\ -\frac{|\dot{u}(t)|}{r(\dot{u}(t))} \end{bmatrix}, \tag{22}$$

$$y(t) = \sigma_0 x(t) + \sigma_1 \dot{x}(t) + \sigma_2 \dot{u}(t), \tag{23}$$

which is a generalized Duhem model of the form (1). Since r given in (20) is not a positively homogeneous function of $\dot{u}$, the LuGre model is not necessarily rate independent. In fact, the input-output maps in Figure 4 show that the LuGre model is rate dependent.

As noted in [30] and [31] the LuGre model has local memory. Thus the hysteresis map $\mathcal{H}_\infty(u, x_0)$ of the LuGre model is independent of x_0.

Maxwell-Slip Model

The Maxwell-slip model [7], [12], [13] shown in Figure 5 has N masses and N springs. For $i = 1, \ldots, N$, the mass m_i with displacement x_i is connected by a stiffness k_i to a common termination point whose displacement is u. Associated with each mass is a displacement deadband of width $\Delta_i > 0$, below which the mass does not move, and above which the mass moves with velocity $\dot{u}$, that is, the inertia of the masses is ignored when the mass is sliding. Hence, $k_i \Delta_i$ is the minimum spring force needed to move the mass m_i. Once the mass m_i begins to move, the spring force remains at $k_i \Delta_i$ for all velocities of the mass. Hence, each mass-spring combination in the Maxwell-slip model is subjected to an equivalent Coulomb friction force $F = k_i \Delta_i$.

We can represent this system as the Duhem model

$$\dot{x}_i(t) = [U(-x_i(t) + u(t) - \Delta_i) \quad 1 - U(-x_i(t)$$
$$+ u(t) + \Delta_i)] \begin{bmatrix} \dot{u}_+(t) \\ \dot{u}_-(t) \end{bmatrix}, \tag{24}$$

$$F(t) = \sum_{i=1}^{N} k_i(-x_i(t) + u(t)), \quad i = 1, \ldots, N, \tag{25}$$

where F is the friction force and

$$U(v) \triangleq \begin{cases} 1, & v \geq 0, \\ 0, & v < 0. \end{cases} \tag{26}$$

The Maxwell-slip model (24), (25) is a generalized Duhem model of the form (1), (2). Since $[\dot{u}_+(t) \ \dot{u}_-(t)]^{\mathrm{T}}$ is positively homogeneous, Proposition 1 implies that (24), (25) is rate independent. Since the step function U is discontinuous, the Maxwell-slip model is discontinuous.

Illustrative input-output maps for the Maxwell-slip model for $N = 1$ and $N = 10$ are shown in Figure 6(a) and (b), respectively.

Consider the ith mass-spring element in the Maxwell-slip model. If the deflection $x_i - u$ of the spring is less than Δ_i, then m_i does not move, that is, $x_i = 0$, and the friction force due to the ith element is given by $F = k_i u$ and thus
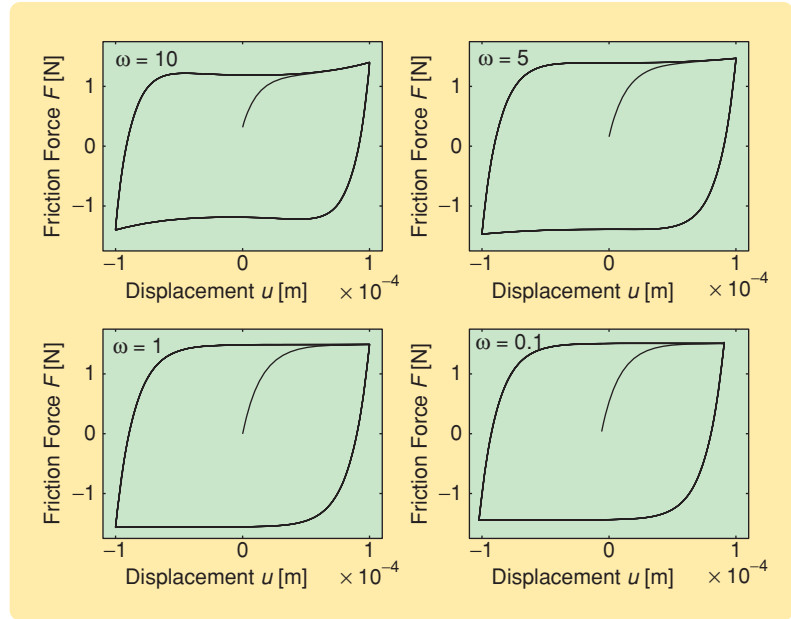

FIGURE 4 Input-output maps for the LuGre model. This sequence of input-output maps at increasingly lower frequencies shows that rate-dependent hysteresis exists from the relative displacement u to the friction force F. The numerical values are $F_{\mathrm{C}} = 1$ N, $F_{\mathrm{S}} = 1.5$ N, $v_{\mathrm{S}} = 0.001$ m/s, $\sigma_0 = 10^5$ N/m, $\sigma_1 = \sqrt{10^5}$ N-s/m, $\sigma_2 = 0.4$ N-s/m, and $u(t) = 10^{-4} \sin \omega t$ m.
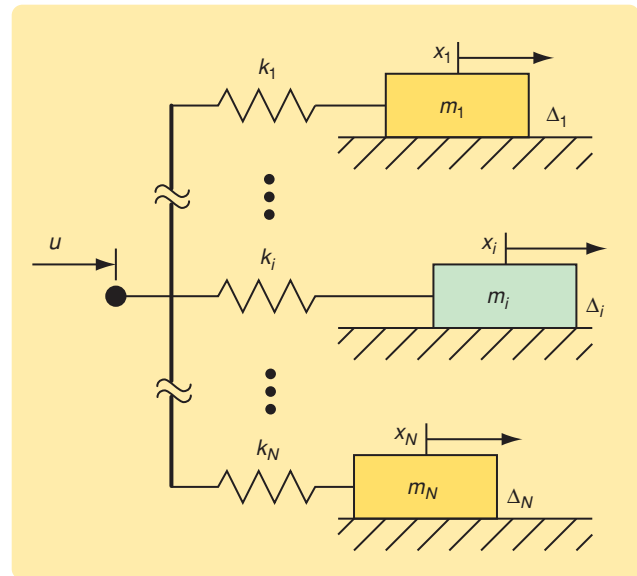

FIGURE 5 The Maxwell-slip model with N masses and N springs. Each mass is associated with a displacement deadband Δ_i, below which the mass does not move and above which the mass moves with the same velocity as the common termination point.

$\dot{F} = k_i \dot{u}$. If the deflection is equal to Δ_i, then $\dot{x}_i = \dot{u}$, and hence $F = k_i \Delta_i$ and thus $\dot{F} = 0$. Consequently, each mass-spring combination in the Maxwell-slip model is a Dahl model with $\gamma = 0$ and $\sigma = k_i$, which has the form

$$\dot{F}(t) = k_i \left[\mathrm{sgn}\left(1 - \frac{F(t)}{F_C}\, \mathrm{sgn}\ \dot{u}(t)\right)\right] \dot{u}(t).$$
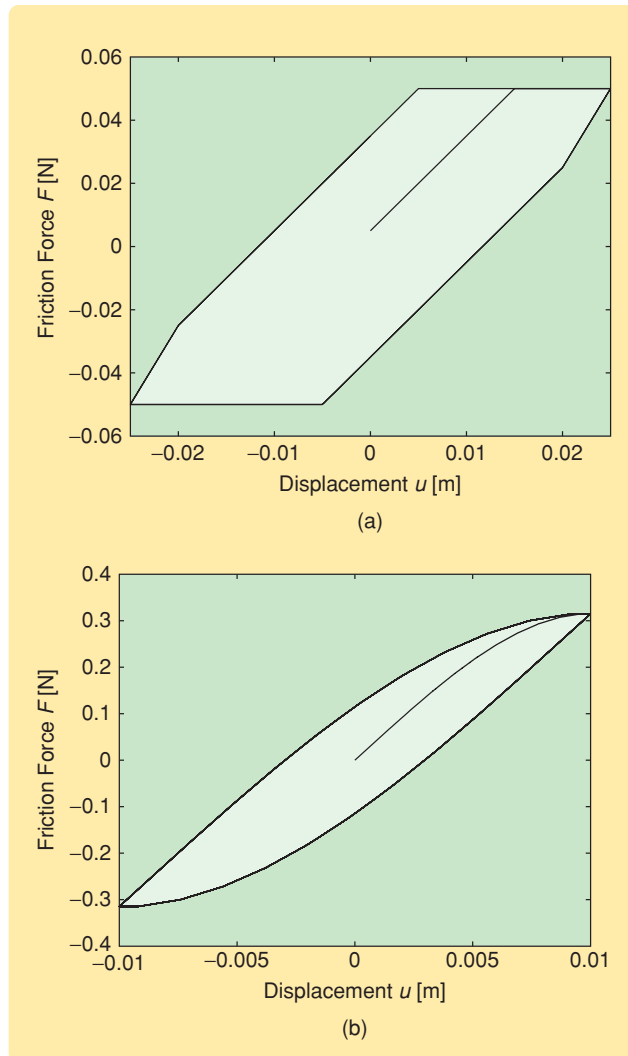
FIGURE 6 Input-output maps for the Maxwell-slip model (a) with $N = 2$, $\Delta_1 = [2.5, 15] \times 10^{-3}$ m, $k = [2, 3]$ N/m, and $u(t) = 0.01 \sin 0.1t$ m, and (b) with $N = 10$, $\Delta = [1.5, 2.4, 3.3, 4.2, 5.1, 6, 6.9, 7.8, 8.7, 9.6]$ $\times 10^{-3}$ m, $k = [1, 1.8, 2.6, 3.4, 4.2, 5, 5.8, 6.6, 7.4, 8.2]$ N/m, and $u(t) = 0.01 \sin 0.1t$ m. The maps show hysteresis from the displacement u to the friction force F.

The Maxwell-slip model has nonlocal memory. Consider (24), (25) with $N = 2$ and $N = 10$ shown in Figure 7(a) and (b), respectively. The input $u(t)$ is initially $u_1(t) = 0.025 \sin(0.1t)$ m in Figure 7(a) and $u_1(t) = 0.01 \sin(0.1t)$ m in Figure 7(b), while the friction force corresponds to the major loops in (a) and (b). When $u(t)$ changes to $u_2(t) = 0.005 \sin(0.1t)$ m after one period, the friction force F corresponds to the upper minor loops in Figure 7(a) and (b). When $u(t)$ changes to $u_2(t)$ after one and a half periods, F corresponds to the lower minor loops in Figure 7(a) and (b). Consequently, with identical inputs but different initial conditions, (24), (25) result in distinct hysteresis maps for the same input $u(t)$. Thus, $\mathcal{H}_\infty(u, x_0)$ depends on x_0, and the Maxwell-slip model has nonlocal memory.

SLIDING BEHAVIOR OF THE FRICTION MODELS

We now consider the behavior of the presliding friction models in the sliding regime, that is, the behavior of these models when subject to large-magnitude displacement and velocity.

Dahl Model

Consider the Dahl model (11) with $\gamma = 1$. The friction force F as a function of displacement u and velocity $\dot{u}$ is shown in Figure 8, where u and $\dot{u}$ are initially set to zero. The position is -20 m when the sign of the velocity changes from negative to positive, and is 0 m when the sign of the velocity changes from positive to negative.

As mentioned above, Figure 8(b) shows that each velocity reversal leads to a delayed change in the sign of the friction force.

LuGre Model

The friction force F as a function of displacement u and velocity $\dot{u}$ for the LuGre model is shown in Figure 9. The Stribeck effect causes the friction force to drop at low magnitudes of velocity.

Maxwell-Slip Model

The friction force F as a function of displacement u and velocity $\dot{u}$ for the Maxwell-slip model is shown in Figure 10. The behavior is similar to that of the Dahl model, that is, sign reversals of the friction force lag velocity sign changes.

ENERGY DISSIPATION DUE TO FRICTION

The energy dissipated during one cycle of motion of a mechanical system in periodic motion is equal to the area enclosed by the counterclockwise force-displacement loop. Since the hysteresis in a friction model occurs between the

We consider three friction models, namely the Dahl model, the LuGre (Lund/Grenoble) model, and the Maxwell-slip model.

friction force and the relative displacement of the surfaces, the area enclosed by the hysteresis map is approximately equal to the kinetic energy lost due to friction at low frequencies. If the hysteresis is rate independent, then the area enclosed by the hysteresis map is equal to the energy dissipated during one cycle of operation at all frequencies.

Dahl Model

Consider the Dahl model (11) with $\gamma = 1$. In this case, the hysteresis map shown in Figure 1 encloses an area of 2.6 N-m = 2.6 J. Since the hysteresis is rate independent, the energy dissipated during one cycle of operation at all frequencies is 2.6 J.

LuGre Model

For the LuGre model (18), (19), the curve in Figure 4 for the input frequency $\omega = 0.1$ rad/s is essentially a hysteresis map. The area enclosed by this curve and thus dissipated during one cycle of operation at low frequency, is approximately 5×10^{-4} J.

Maxwell-Slip Model

Consider the Maxwell-slip model with N mass-spring elements as shown in Figure 5. Under periodic motion, the springs gain potential energy in compression and extension but do not return this energy to the system. The lost energy is therefore the energy dissipated by the friction. Provided $u_{max} > \Delta_i$ for all i, where u_{max} is the maximum value of the periodic input, the potential energy gained by the springs in one complete cycle of the periodic input is given by $E_{diss} = \sum_{i=1}^{N} k_i \Delta_i^2$. For the Maxwell slip model with ten elements considered in Figure 6(b), $E_{diss} = 2.4 \times 10^{-3}$ J, where J denotes Joule. By numerical integration, the area enclosed by the hysteresis map in Figure 6(b) is found to be approximately 2.5×10^{-3} J.

HYSTERESIS INDUCED BY DUHEM FEEDBACK

To study hysteresis induced by Duhem friction models, we consider the feedback interconnection of a single-input, single-output linear system and a Duhem hysteretic model as shown in Figure 11.

The following definition given in [32] is needed.

Definition 4

Consider the system in Figure 11 with constant u. The system is step convergent if $\lim_{t \to \infty} y(t)$ exists for all initial conditions and for all $u \in \mathbb{R}$.

Suppose the feedback system in Figure 11 is step convergent. Then it follows from Definition 4 that $\lim_{t \to \infty} y(t)$ exists for every constant u. Now, let $u : [0, \infty) \to [u_{min}, u_{max}]$ be periodic with period α, and let $u_T(t) = u(\alpha t/T)$ for all t. Now assume that, for all $x_0 \in \mathbb{R}^n$, the periodic input-output map $\mathcal{H}_T(u_T, y_T, x_0)$ exists for all $T > 0$ and that the limiting periodic input-output map
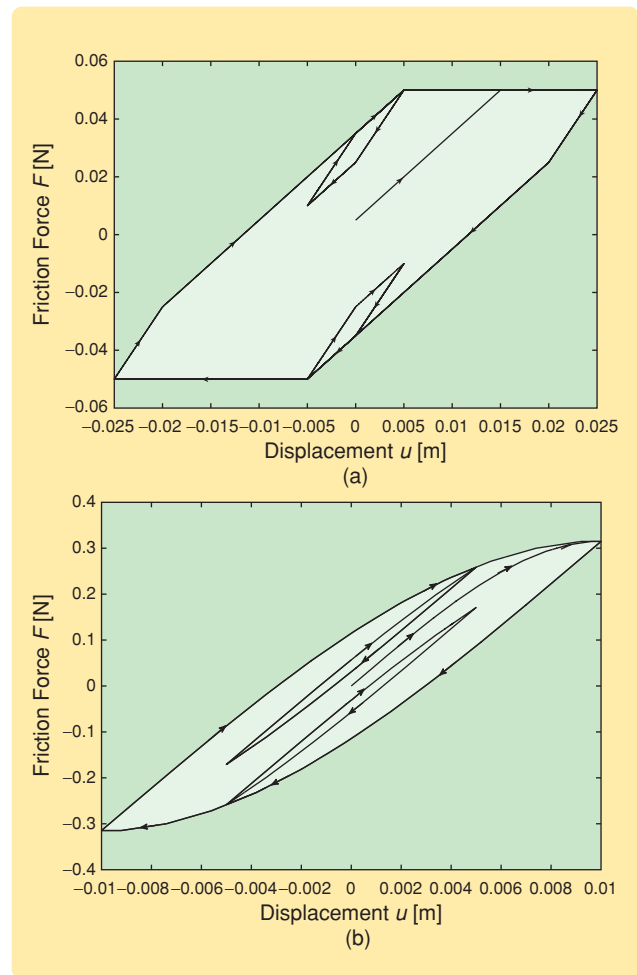

FIGURE 7 Nonlocal memory of the Maxwell-slip model with (a) $N = 2$, $\Delta_1 = 2.5 \times 10^{-3}$ m, $\Delta_2 = 1.5 \times 10^{-2}$ m, $k_1 = 2$ N/m, and $k_2 = 3$ N/m, and (b) with $N = 10$, $\Delta = [1.5, 2.4, 3.3, 4.2, 5.1, 6, 6.9, 7.8, 8.7, 9.6] \times 10^{-3}$ m, and $k = [1, 1.8, 2.6, 3.4, 4.2, 5, 5.8, 6.6, 7.4, 8.2]$ N/m. The input $u(t)$ is initially $u_1(t) = 0.025 \sin(0.1t)$ m in (a) and $u_1(t) = 0.01 \sin(0.1t)$ m in (b), where the friction force corresponds to the major loops. When $u(t)$ changes to $u_2(t) = 0.005 \sin(0.1t)$ m after one period, the friction force F corresponds to the upper minor loops. When $u(t)$ changes to $u_2(t)$ after one and a half periods, the friction force F corresponds to the lower minor loops.

$\mathcal{H}_\infty(u, x_0)$ exists. Since the set $\mathcal{H}_\infty(u, x_0)$ represents the response of the feedback system in the limit of dc operation, each element (u, y) of $\mathcal{H}_\infty(u, x_0)$ is the limit of a sequence of points $\{(u_i, y_i)\}_{i=1}^\infty$, where $(u_i, y_i) \in \mathcal{H}_{T_i}(u_{T_i}, y_{T_i}, x_0)$, as $T_i \to \infty$, that is, as the input becomes increasingly slower. Since a constant input u_∞ can be viewed as a periodic input with infinite period, the component y_∞ of each limiting point $(u_\infty, y_\infty) \in \mathcal{H}_\infty(u, x_0)$ is given by $y_\infty = \lim_{t \to \infty} y(t)$ under the constant input u_∞. This observation suggests that step convergence of the feedback system is necessary and sufficient for the existence of $\mathcal{H}_\infty(u, x_0)$.

Let $\bar{y}_1$ and $\bar{z}_1$ represent limiting values of the outputs of the linear system $G(s)$ and the Duhem model, respectively, for a constant input u. If $G(s)$ has no poles in the closed right-half plane and the system in Figure 11 is step convergent, then

$$\bar{y}_1 = G(0)(u - \bar{z}_1). \tag{27}$$

Now suppose that, for some constant u, the hysteretic Duhem model has distinct equilibria $\bar{z}_1$ and $\bar{z}_2$. Then the output of the feedback system also has distinct equilibria $\bar{y}_1 = G(0)(u - \bar{z}_1)$ and $\bar{y}_2 = G(0)(u - \bar{z}_2)$. Hence, the limiting periodic map $\mathcal{H}_\infty(u, x_0)$ exists, that is, there exists hysteresis between $u(t)$ and $y(t)$. Furthermore, for a given u, the horizontal width of the hysteresis map is given by $\bar{y}_1 - \bar{y}_2 = G(0)(\bar{z}_2 - \bar{z}_1)$. In the following section we illustrate these observations with an example.

HYSTERESIS INDUCED BY FRICTION IN A MASS-SPRING SYSTEM

We now consider the force-actuated mass-spring system shown in Figure 12, whose dynamics are given by
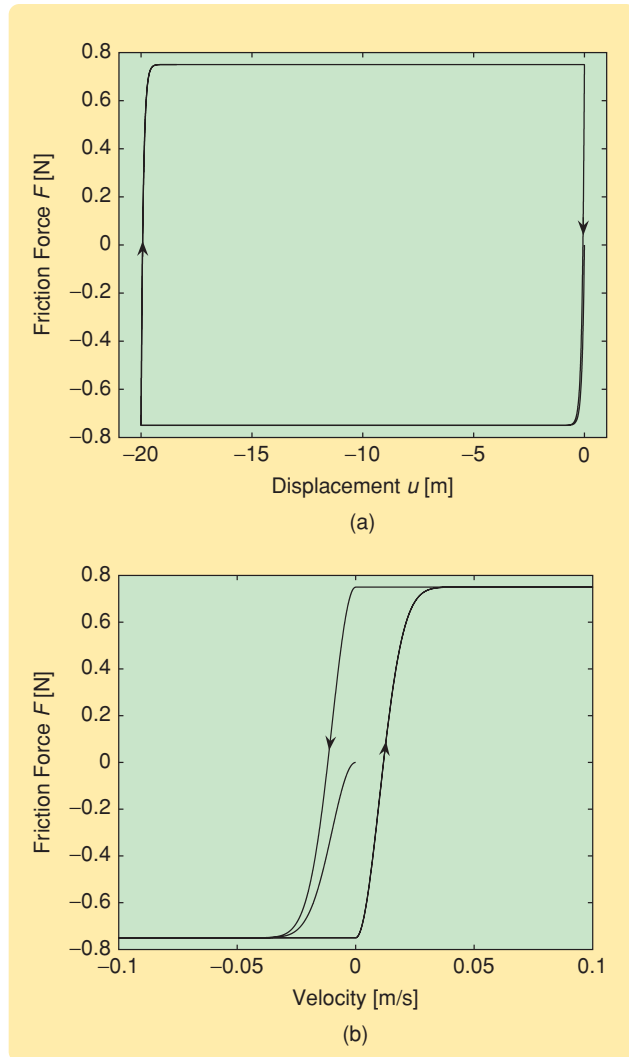

FIGURE 8 (a) Friction force versus displacement and the corresponding (b) friction force versus velocity for the Dahl model with $F_C = 0.75$ N, $\gamma = 1$, $\sigma = 7.5$ N/m, and $u(t) = 10((\cos 0.01 t) - 1)$ m. The friction force saturates at F_C, while its sign reversal lags the sign change in velocity.
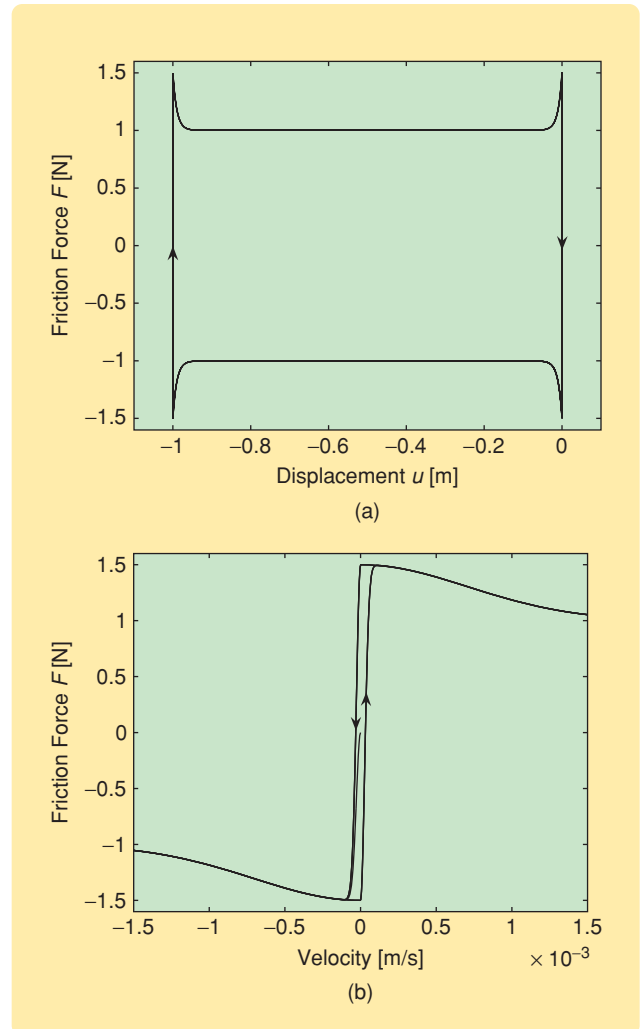

FIGURE 9 (a) Friction force versus displacement and the corresponding (b) friction force versus velocity for the LuGre model (18), (19) with $F_C = 1$ N, $F_S = 1.5$ N, $v_S = 0.001$ m/s, $\sigma_0 = 10^5$ N/m, $\sigma_1 = \sqrt{10^5}$ N-s/m, $\sigma_2 = 0.4$ N-s/m, and $u(t) = 0.01((\cos 0.01 t) - 1)$ m. The friction force saturates at F_S, while its sign reversal lags the sign change in velocity.

$$\ddot{q}(t) + \frac{k_s}{m}q(t) = \frac{1}{m}F_e(t) - \frac{1}{m}F, \qquad (28)$$

where $q(t)$ is the displacement of the mass, k_s is the spring constant, m is the mass, $F_e(t)$ is the external force exerted on the mass, and F is the friction force acting on the mass. Let $v = \dot{q}$ denote the velocity of the mass.

In Figure 11, this system is represented as the feedback interconnection of a linear system with a Duhem hysteretic model, where $u = F_e$, $y = q$, $z = F$, and the transfer function $G(s)$ is given by

$$G(s) = \frac{1}{ms^2 + k_s}. \qquad (29)$$

Dahl Model

Using the Dahl model (11) for the friction force with mass displacement $u = q$, we have

$$\dot{F}(t) = \sigma \left| 1 - \frac{F(t)}{F_C}\mathrm{sgn}\,\dot{q}(t) \right|^{\gamma} \mathrm{sgn}\left(1 - \frac{F(t)}{F_C}\mathrm{sgn}\,\dot{q}(t)\right)\dot{q}(t). \qquad (30)$$

As shown above, the friction force given by the Dahl model converges to the constant values F_C or $-F_C$ depending on the sign of the relative velocity and acts in a direction opposing the motion. Consequently, the Dahl friction acting on the mass in the mass-spring system plays the role of a damper. Hence, the force-actuated mass-spring system with Dahl friction is step convergent, and the states in the feedback representation (29), (30) converge to constant values for every constant $F_e \in \mathbb{R}$. Letting $\bar{q}$, $\bar{v}$, and $\bar{F}$ denote steady-state values, (27), (28), and (30) yield
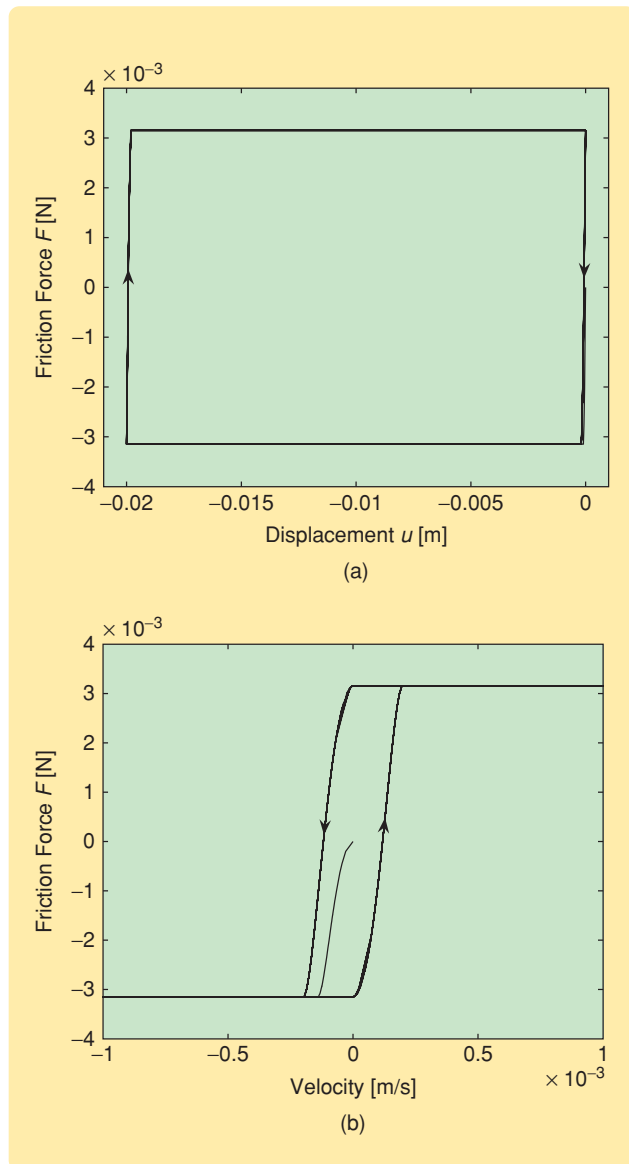

FIGURE 10 (a) Friction force versus displacement and the corresponding (b) friction force versus velocity for the Maxwell-slip model with $N = 10$, $\Delta = [1.5, 2.4, 3.3, 4.2, 5.1, 6, 6.9, 7.8, 8.7, 9.6] \times 10^{-5}$ m, $k = [1, 1.8, 2.6, 3.4, 4.2, 5, 5.8, 6.6, 7.4, 8.2]$ N/m, and $u(t) = 0.01((\cos 0.01t) - 1)$ m. The friction force sign reversal lags the sign change in velocity.
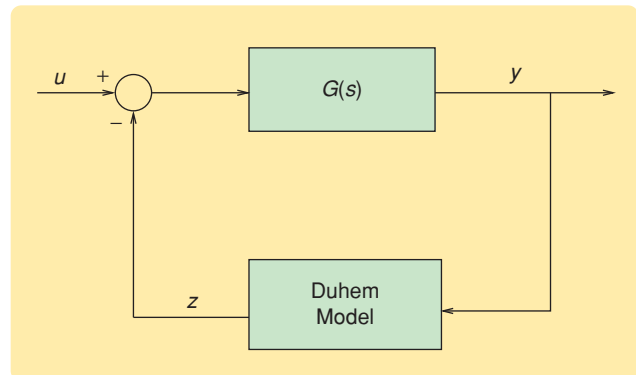

FIGURE 11 Single-input, single-output linear system with Duhem feedback. This model is used to study hysteresis induced by a Duhem friction model when connected by feedback to a linear system.
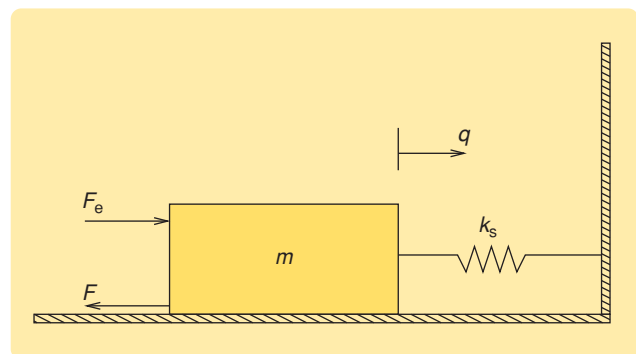

FIGURE 12 Force-actuated mass-spring system. The friction force is denoted by F, while the external force is F_e. The system exhibits hysteresis from the external force F_e to the displacement q when the Dahl and LuGre models are used to model the friction force F.

$$\bar{v} = 0, \tag{31}$$

$$\bar{q} = G(0)(u - \bar{z}) = \frac{F_e - \bar{F}}{k_s}, \tag{32}$$

$$\left(1 - \frac{\bar{F}}{F_C}\operatorname{sgn}\bar{v}\right) = 0. \tag{33}$$

For a constant external force input, the steady-state values of displacement, velocity, and friction force are given by (31), (32), and (33), respectively. Thus, for a low-frequency external force input, $v(t) \to 0$ as $t \to \infty$, and the displacement of the mass satisfies $q(t) \to \bar{q}$ as $t \to \infty$, where $\bar{q}$ is given by (32). Now, if (33) holds, then $\bar{F} = F_C$ or $\bar{F} = -F_C$ depending on the sign of $v(t)$. Consequently, from (32), the steady-state displacement $\bar{q}$ can assume two different values, namely,

$$\bar{q}_1 = G(0)(F_e + F_C) = \frac{F_e + F_C}{k_s},$$

$$\bar{q}_2 = G(0)(F_e - F_C) = \frac{F_e - F_C}{k_s}.$$

These observations suggest that the limit of the periodic map $\mathcal{H}_T(F_{eT}, q_T)$ exists as $T \to \infty$, that is, there exists hysteresis between $F_e(t)$ and $q(t)$. The width of the map is given by $\bar{q}_1 - \bar{q}_2 = 2F_c/k_s$. For $k_s = 1.5$ N/m, $F_c = 0.75$ N,

and $F_e(t) = 5\sin(0.001t)$ N, the hysteresis map from the mass displacement $q(t)$ to the external force $F_e(t)$ is shown in Figure 13. It can be seen that the vertical width of the hysteresis map is 1 m.

LuGre Model

Using the LuGre model (18), (19) with $u(t) = q(t)$, we have

$$\dot{x}(t) = \dot{q}(t) - \frac{|\dot{q}(t)|}{r(\dot{q}(t))}x(t), \tag{34}$$

$$F(t) = \sigma_0 x(t) + \sigma_1 \dot{x}(t) + \sigma_2 \dot{q}(t), \tag{35}$$

where

$$r(\dot{q}(t)) = \frac{F_C}{\sigma_0} + \frac{F_S - F_C}{\sigma_0}e^{-(\dot{q}(t)/v_S)^2}. \tag{36}$$

Due to the Stribeck effect, relating the equilibria map to the hysteresis map in the case of the LuGre model is more complicated compared to the Dahl model. The friction force acts in a direction opposing the motion and, consequently, plays the role of a damper in the force-actuated mass-spring system. Hence, the force-actuated
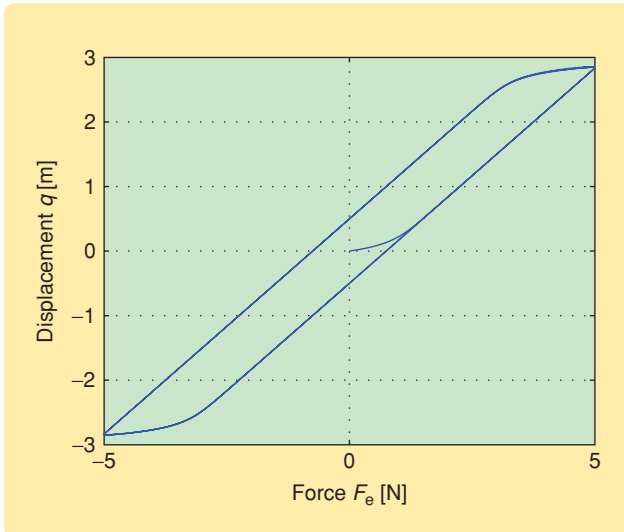

FIGURE 13 Hysteresis map from the external force $F_e(t)$ to the displacement $q(t)$ of the mass for the force-actuated mass-spring system (28) in Figure 12, with the Dahl model. The numerical values are $F_C = 0.75$ N, $\gamma = 1$, $\sigma = 7.5$ N/m, $k_s = 1.5$ N/m, $m = 1$ kg, and $F_e(t) = 5\sin(0.001t)$ N. The vertical width of the hysteresis map is 1 m.
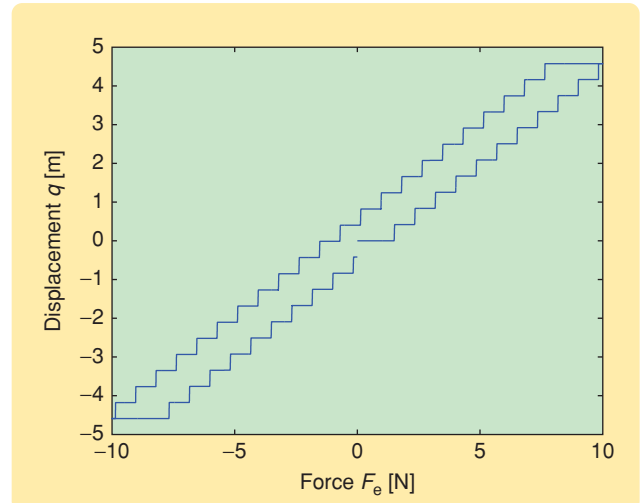
FIGURE 14 Hysteresis map from the external force $F_e(t)$ to the displacement $q(t)$ of the mass for the force-actuated mass-spring system (28) in Figure 12, with the LuGre model (34), (35). The staircase pattern is caused by the Stribeck effect in the LuGre model. The numerical values are $F_C = 1$ N, $F_S = 1.5$ N, $v_S = 0.001$ m/s, $\sigma_0 = 10^5$ N/m, $\sigma_1 = \sqrt{10^5}$ N-s/m, $\sigma_2 = 0.4$ N-s/m, and $F_e(t) = 10\sin(0.001t)$ N.

mass-spring system with LuGre friction is step convergent, and the states in the feedback representation given by (29), (34), and (35) converge to constant values for every constant $F_e \in \mathbb{R}$.

The hysteresis map from the input F_e to the output q for a low-frequency input $F_e = 10\sin(0.001t)$ N is shown in Figure 14. The time histories of the friction force F and the position of the mass q are shown in Figure 15. The mass-spring system exhibits stick-slip motion [10] in which the mass sticks until the friction force exceeds the breakaway force F_S. Once the mass starts moving, the friction force drops because of the Stribeck effect. Consequently, the mass accelerates, and thus the spring contracts and the spring force increases. The mass accelerates until the external force is balanced by the friction force and the spring force. When the spring force becomes sufficiently large, the mass decelerates and sticks again. This process repeats. Stick-slip is reflected by the regions of zero velocity shown in Figure 15(b). The staircase pattern in the hysteresis map shown in Figure 14 is caused by the stick-slip motion. The low-slope horizontal segments of the map correspond to sticking, while the high-slope vertical segments correspond to slip. It should be noted that the hysteresis map is continuous despite the steep vertical segments.

The amplitude of the oscillations in the friction force shown in Figure 15(a) is equal to $F_S - F_C = 0.5$ N. The length ΔF_e of the horizontal segments of the hysteresis map is twice the amplitude of the oscillations in the friction force, that is, $\Delta F_e = 1$ N. The horizontal segments correspond to the sticking phase of the motion, in which the mass is at rest and thus the external force is balanced by the spring force and the friction force. As the external force increases, the friction force also increases until reaching F_S, after which the mass slips, the friction force drops, and the spring force increases. The larger spring force causes the mass to stick again, leading to new balanced forces.

The vertical segments of the hysteresis map correspond to the slipping phase, and their size can be determined by balancing forces. For instance, consider the first vertical step starting from the origin in the hysteresis map in Figure 15(b). Let Δq denote the length of the vertical segment. The external force is $F_e = 1.5$ N, and it can be seen in Figure 15(a) that the friction force drops to $F = 0.6$ N. By balancing the forces, we have thus the spring force

$$k_s \Delta q = F_e - F = 1.5 - 0.6 = 0.9 \ \text{N},$$

which implies that
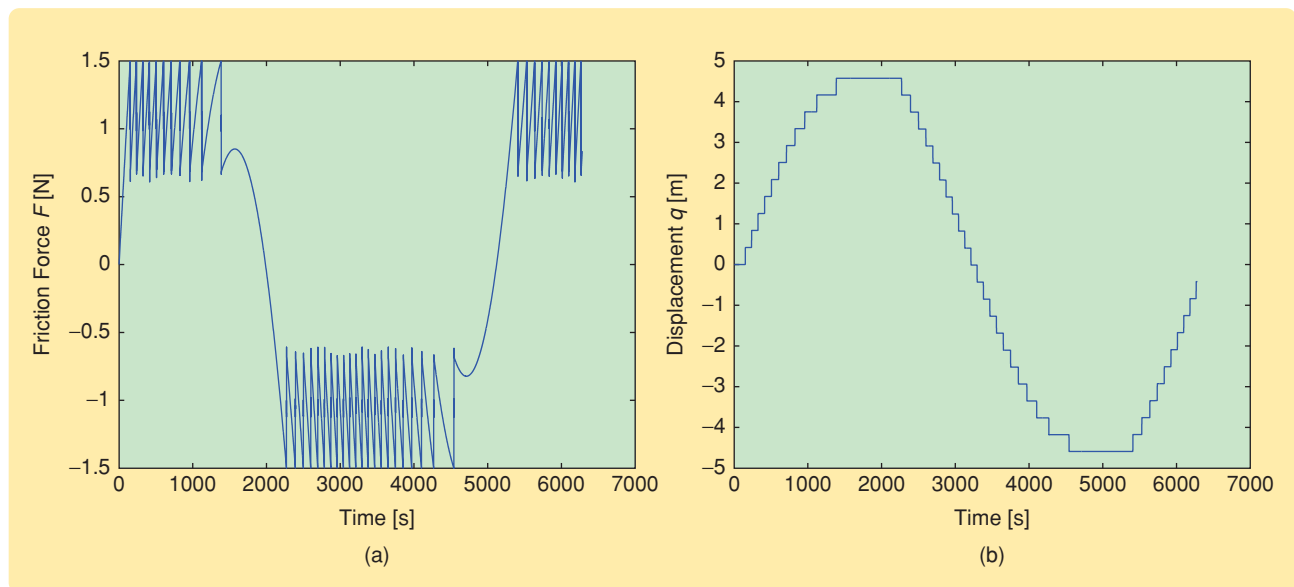
FIGURE 15 (a) Friction force $F(t)$ and (b) displacement $q(t)$ of the mass for the force-actuated mass-spring system (28) in Figure 12, with the LuGre model (34), (35). The friction force oscillates with amplitude $F_S - F_C$, and the mass exhibits stick-slip motion. The numerical values are $F_C = 1$ N, $F_S = 1.5$ N, $v_S = 0.001$ m/s, $\sigma_0 = 10^5$ N/m, $\sigma_1 = \sqrt{10^5}$ N-s/m, $\sigma_2 = 0.4$ N-s/m, and $F_e(t) = 10\sin(0.001t)$ N.

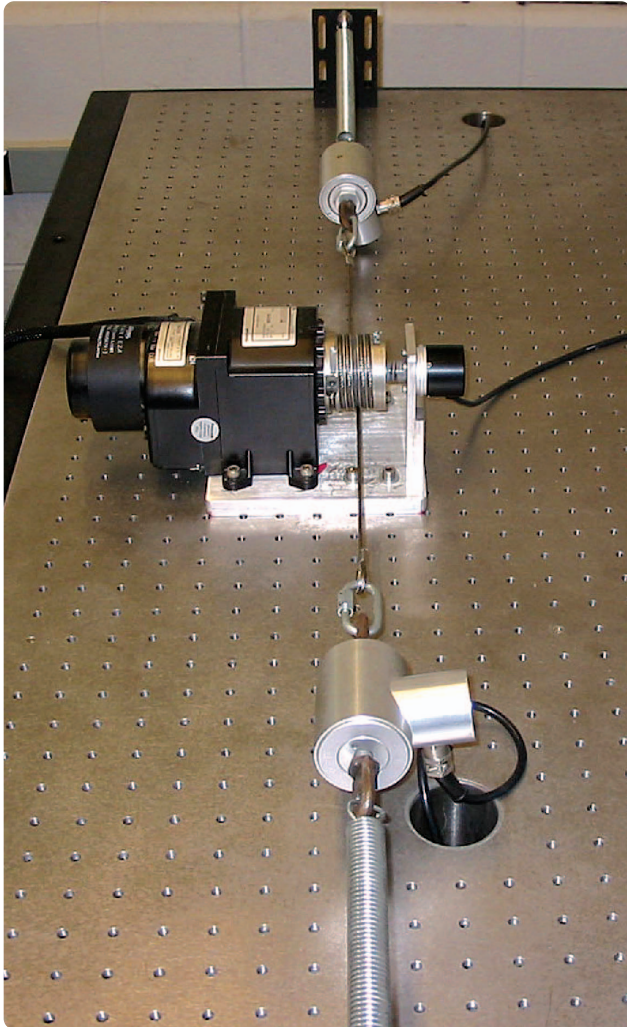$$\Delta q = \frac{0.9}{k_s} = \frac{0.9}{2} \text{ m} = 0.45 \text{ m}.$$

FIGURE 16 Experimental setup. This testbed involves a dc motor with a gearbox for studying the effects of gearbox friction. A pair of cables are wound around a drum attached to the gearbox and connected to a pair of springs through two load cells. The forces exerted by the springs are measured by the load cells.
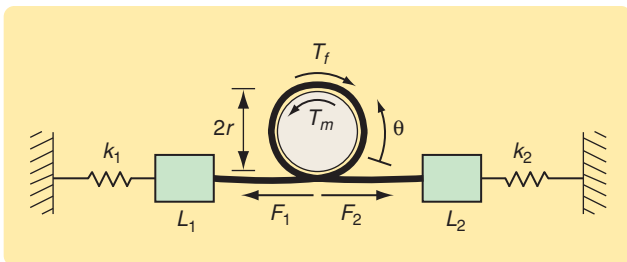

FIGURE 17 Schematic of the experimental setup looking toward the drum along its axis. The motor is behind the drum and is not shown. The drum angle θ is positive for counterclockwise rotation of the shaft.

The hysteresis map can thus be completely determined in terms of the parameters F_S, F_C, m, and k_s.

EXPERIMENTAL SETUP

We now describe an experimental setup for studying the effects of gearbox friction on the dynamics of a dc motor. The experimental setup is shown in Figure 16, while a schematic of the setup is shown in Figure 17. Two cables are wound around a drum attached to the motor shaft and connected to load cells L_1 and L_2, which measure the force exerted by the springs k_1 and k_2. Since hysteresis is an asymptotically low-frequency phenomenon, the inertia of the load cells does not affect the hysteretic properties of the system.

The dynamics of the drum are given by

$$I\ddot{\theta} = T_m - T_f + F_2 r - F_1 r, \tag{37}$$

where θ is the angle of rotation of the drum, I is the drum's moment of inertia, T_m is the torque exerted by the motor, T_f is the torque due to friction, r is the radius of the drum, and F_1 and F_2 are the forces exerted on the drum by the springs. The cables are wound such that, as the shaft rotates counterclockwise, F_1 increases and F_2 decreases. The springs are sufficiently pre-stressed that neither spring slacks while the shaft rotates in either direction. Let F_{10} and F_{20} denote the values of the spring forces at the initial drum angle, and let δ_1 and δ_2 denote the deflections in the two springs. Then

$$F_1 = F_{10} + k_1 \delta_1, \quad F_2 = F_{20} + k_2 \delta_2$$

and

$$\delta_1 = r\theta, \quad \delta_2 = -r\theta.$$

Hence,

$$I\ddot{\theta} = T_m - T_f - (F_{10} - F_{20})r - (k_1 + k_2)r^2\theta. \tag{38}$$

The motor torque T_m is assumed to be proportional to the motor current, that is,

$$T_m = k_m i_m, \tag{39}$$

where k_m is the proportionality constant and i_m is the motor current. Hence, (38) becomes

$$\ddot{\theta}(t) + \frac{(k_1 + k_2)r^2}{I}\theta(t) = \frac{k_m}{I}i_m(t) - \frac{1}{I}T_f - \frac{(F_{10} - F_{20})r}{I}. \tag{40}$$

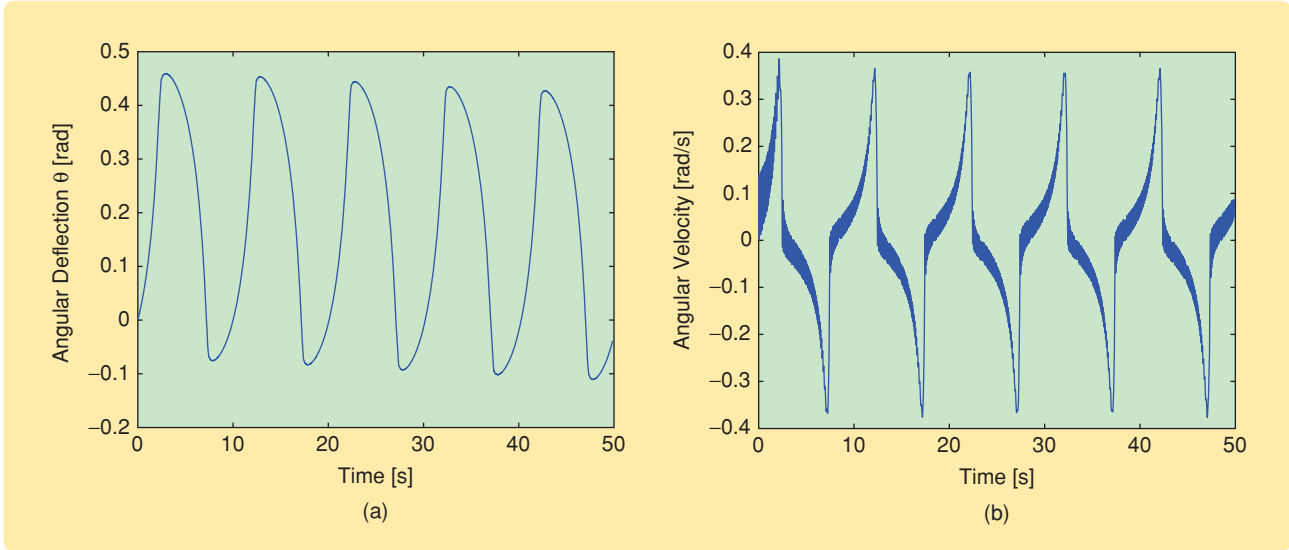If $F_{10} = F_{20}$, then, with the correspondences

FIGURE 18 Simulation of the experimental setup with the Dahl model. (a) Drum angle θ and (b) drum angular velocity $\dot{\theta}$. The numerical values are $F_C = 0.84$ N-m, $\gamma = 1$, $\sigma = 7.5$ N-m/rad, and $i_m(t) = 0.05 \sin(0.2\pi t)$ A.

$$\frac{k_s}{m} = \frac{(k_1 + k_2)r^2}{I},$$
$$\frac{F_e}{m} = \frac{k_m}{I}i_m,$$
$$\frac{F}{m} = \frac{1}{I}T_f, \tag{41}$$

the dynamics in (40) are identical to the dynamics of the mass-spring system (28).

The setup is connected to a digital computer through a dSPACE 1103 system, which has one encoder, five analog to digital (A/D) channels, and five digital to analog (D/A) channels. Each load cell, whose output is amplified by an Endevco voltage amplifier model 136, can measure a maximum load of 75 kg and has a sensitivity of 0.26 mV/kg. The amplifier gain can be set between zero and 1000, and the amplified signals are sampled by the dSPACE system. The dc motor has a built-in tachometer that measures the angular velocity of the motor shaft. The angular velocity signal is read through an A/D channel. The conversion for the tachometer output is 0.01 V/rpm. A Heidenhain encoder measures the angle of the drum. The gear ratio between the motor shaft and the drum is 1:68.8. Current is supplied to the dc motor through a Quanser linear current amplifier LCAM. The required current profile is commanded to the current amplifier through one of the D/A channels. The

amplifier provides a voltage signal proportional to the current supplied to the dc motor. Estimated parameter values are $k_1 = k_2 = 2$ N/m, $F_{10} = F_{20} = 0.01$ N, $r = 1$ inch, $k_m = 16.5$ N-m/A, and $I = 3 \times 10^{-4}$ kg-m².

FIGURE 19 Simulation of hysteresis in the experimental setup with the Dahl model. The hysteresis map is from the motor torque T_m to the drum angle θ. The numerical values are $F_C = 0.84$ N-m, $\gamma = 1$, $\sigma = 7.5$ N-m/rad, and $i_m(t) = 0.05 \sin(0.2\pi t)$ A.
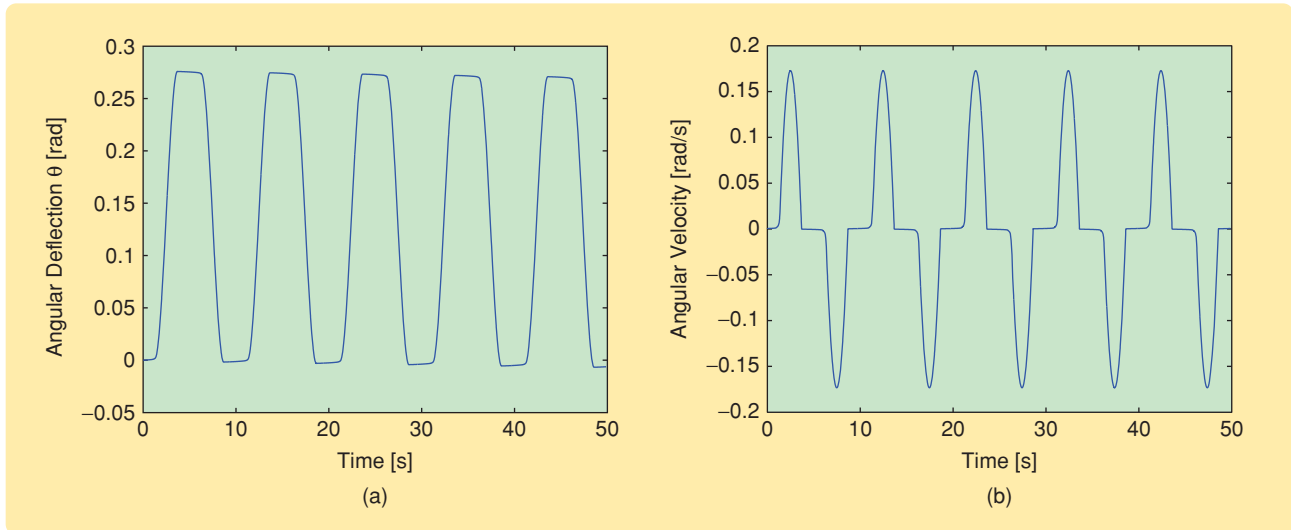
FIGURE 20 Simulation of the experimental setup with the LuGre friction model showing (a) drum angle θ and (b) drum angular velocity. The numerical values are $F_C = 0.6$ N-m, $F_S = 1$ N-m, $v_S = 0.001$ rad/s, $\sigma_0 = 10^3$ N-m/rad, $\sigma_1 = \sqrt{10^3}$ N-m-s/rad, $\sigma_2 = 1.3$ N-m-s/rad, and $i_m(t) = 0.05\sin(0.2\pi t)$ A.

SIMULATION RESULTS

In this section we simulate the dc motor dynamics (40). The friction torque T_f is modeled using the Dahl, LuGre, and Maxwell-slip models. The motor current is chosen to be $i_m(t) = 0.05\sin(0.2\pi t)$ A. For the Dahl model, the drum angle θ and the drum angular velocity $\dot{\theta}$ are shown in Figure 18. The hysteresis map from the motor torque T_m to the drum angle θ is shown in Figure 19. For the LuGre model, θ and $\dot{\theta}$ are shown in Figure 20, while the hysteresis map from the motor torque to the drum angle θ is shown in Figure 21. For the Maxwell-slip model, θ and $\dot{\theta}$ are shown in Figure 22, while the hysteresis map from the motor torque to the drum angle θ is shown in Figure 23.
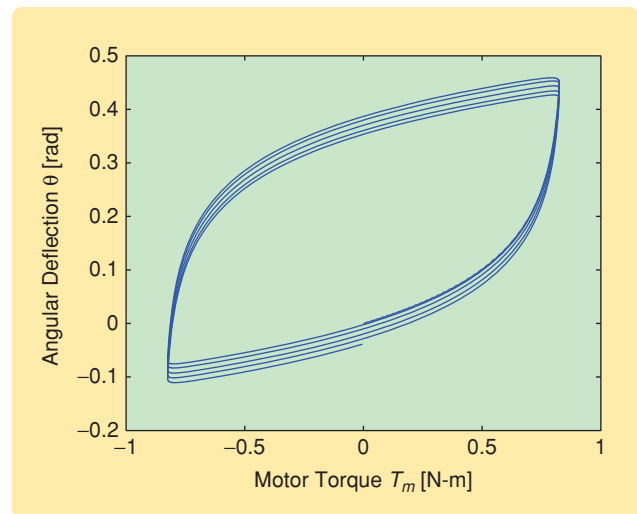

FIGURE 21 Simulation of hysteresis in the experimental setup with the LuGre model. The hysteresis map is from the motor torque T_m to the drum angle θ. The numerical values are $F_C = 0.6$ N-m, $F_S = 1$ N-m, $v_S = 0.001$ rad/s, $\sigma_0 = 10^3$ N-m/rad, $\sigma_1 = \sqrt{10^3}$ N-m-s/rad, $\sigma_2 = 1.3$ N-m-s/rad, and $i_m(t) = 0.05\sin(0.2\pi t)$ A.
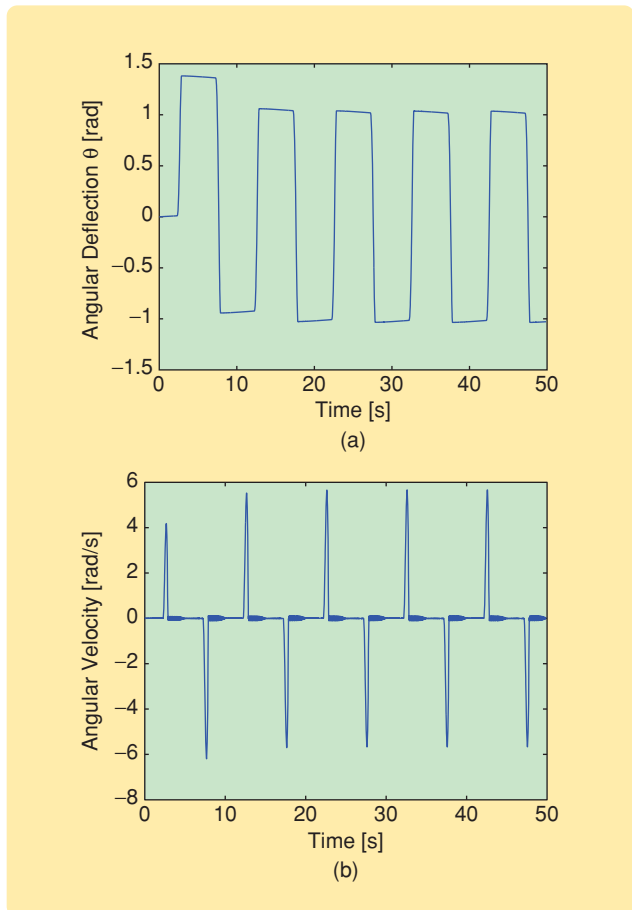
FIGURE 22 Simulation of the experimental setup with the Maxwell-slip model. (a) Angular deflection θ and (b) angular velocity $\dot{\theta}$ of the motor shaft. The numerical values are $N = 10$, $\Delta = [1.5,\ 2.4,\ 3.3,\ 4.2,\ 5.1,\ 6,\ 6.9,\ 7.8,\ 8.7,\ 9.6] \times 10^{-3}$ rad, $k = [2.60,\ 4.68,\ 6.76,\ 8.84,\ 10.92,\ 13.00,\ 15.08,\ 17.16,\ 19.24,\ 21.32]$ N-m/rad, and $i_m = 0.05\sin(0.2\pi t)$ A.

EXPERIMENTAL RESULTS

In this section we present and analyze results obtained from the experimental setup. The current profile commanded to the current amplifier is $0.05\sin(0.2\pi t)$ A. The load cell readings, the actual current supplied by the amplifier to the dc motor, the shaft angular velocity measured by the tachometer, and the drum angle measured by the encoder are shown in Figure 24. Because of asymmetry in the setup, the force values F_{10} and F_{20} in the two springs are not equal. Once the current supply is switched on, the motor shaft reaches an equilibrium position at approximately $t = 5$ s in which both springs are equally stressed and then oscillates about that position. This behavior is evident in the load cell readings as shown in Figure 24.

The motor shaft rotates slowly in one direction until the motor torque cannot exceed the torque exerted by the springs and the gearbox friction, after which the motor stops rotating and remains motionless until the motor torque changes direction. This behavior can be seen in Figure 24 where the shaft has zero angular velocity in periodic time intervals, which matches best with the angular velocity obtained from the simulations using the LuGre model shown in Figure 20(b). The motor torque and drum angle exhibiting hysteresis in the experiment are shown in Figure 25. After initial transients, the motion approaches a hysteresis map. Aside from the transients and the bias in the motor torque, the hysteresis map in Figure 25 has the same form as the map simulated using the LuGre model shown in Figure 21. Consequently, of the three friction models considered, the LuGre model provides the best model of the gearbox friction.

Parameter Identification

Since simulation with the LuGre model gives a hysteresis map that qualitatively resembles the experimental hysteresis map, we determine LuGre friction parameters to quantitatively match the simulated and experimental hysteresis maps. Using a manual tuning approach, suitable parameter values are found to be $F_C = 0.6$ N-m, $F_S = 1$ N-m, $v_S = 0.001$ rad/s, $\sigma_0 = 10^3$ N-m/rad, $\sigma_1 = \sqrt{10^3}$ N-m-s/rad, and $\sigma_2 = 1.3$ N-m-s/rad. Simulation of the experimental setup using these parameters yields

the plots of θ and $\dot\theta$ shown in Figure 20. Figure 26 compares the experimental and simulated hysteresis maps. Adjustments are made for bias in the motor current and asymmetry in the motor setup.
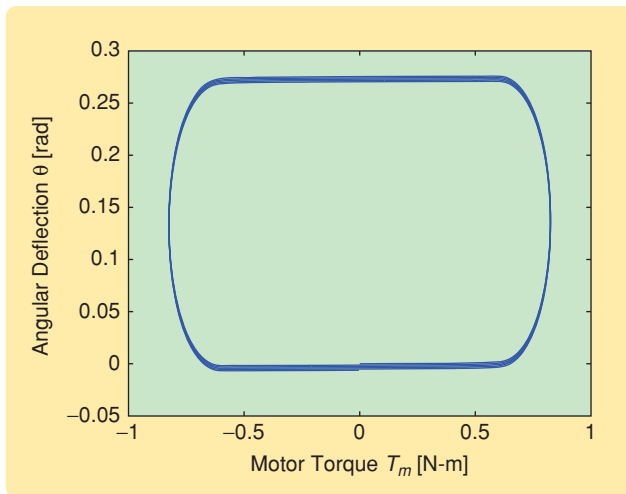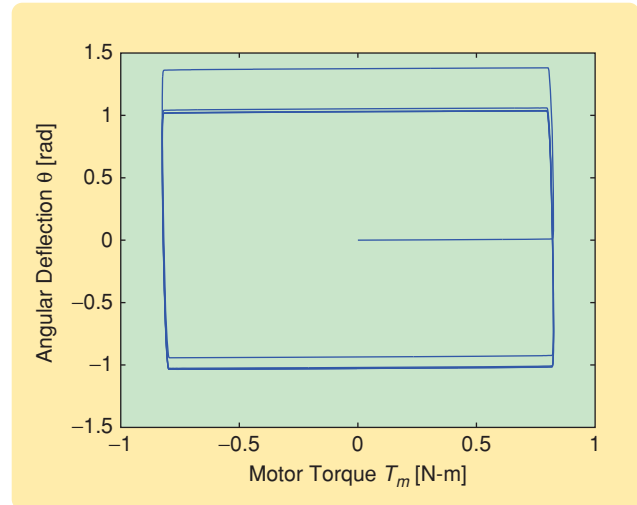
FIGURE 23 Simulation of hysteresis in the experimental setup with the Maxwell-slip model. This hysteresis map is from the motor torque T_m to the drum angle θ. The numerical values are $N = 10$, $\Delta = [1.5, 2.4, 3.3, 4.2, 5.1, 6, 6.9, 7.8, 8.7, 9.6] \times 10^{-3}$ rad, $k = [2.60, 4.68, 6.76, 8.84, 10.92, 13.00, 15.08, 17.16, 19.24, 21.32]$ N-m/rad, and $i_m = 0.05\sin(0.2\pi t)$ A.
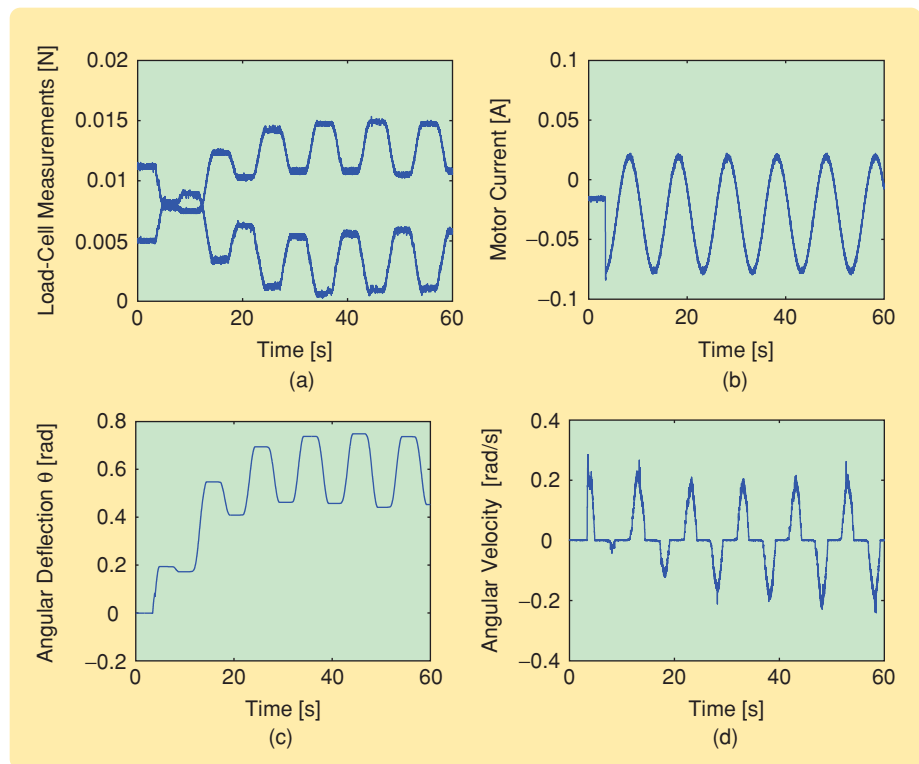
FIGURE 24 Experimental results. These plots show (a) load-cell measurements, (b) current supplied by the amplifier, (c) the drum angle, and (d) the drum angular velocity readings for the dc motor experiment. The input current is $i_m(t) = 0.05\sin(0.2\pi t)$ A. Note that the drum angular velocity plot resembles the plot in Figure 20(b).

FIGURE 25 Hysteresis in the dc motor experiment from the motor torque to the drum angle. The input current is $i_m(t) = 0.05\sin(0.2\pi t)$ A. Aside from the transients and the bias in the motor torque, the hysteresis map is similar in form to the simulated hysteresis map in Figure 21, based on the LuGre model.
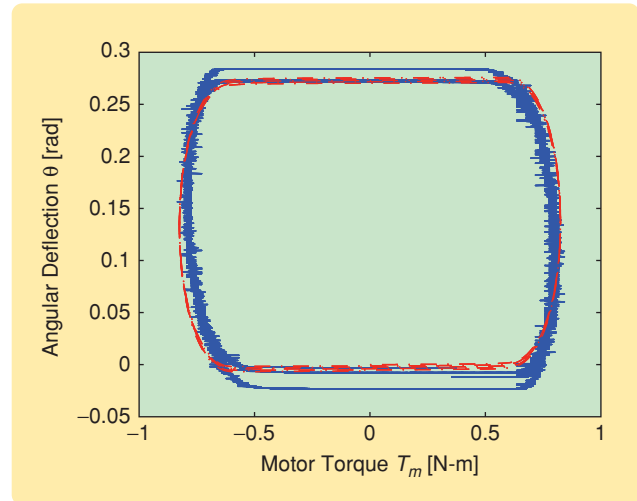
**FIGURE 26** Matching the experimental hysteresis map and the simulated hysteresis map based on the LuGre model. The hysteresis map is from the motor torque T_m to the drum angle θ. The numerical values for the simulations are $F_C = 0.6$ N-m, $F_S = 1$ N-m, $v_S = 0.001$ rad/s, $\sigma_0 = 10^3$ N-m/rad, $\sigma_1 = \sqrt{10^3}$ N-m-s/rad, and $\sigma_2 = 1.3$ N-m-s/rad. The input current is $i_m(t) = 0.05\sin(0.2\pi t)$ A for both the experiment and the simulation.

CONCLUSIONS

In this article we recast the Dahl, LuGre, and Maxwell-slip models as extended, generalized, or semilinear Duhem models. We classified each model as either rate independent or rate dependent. Smoothness properties of the three friction models were also considered.

We then studied the hysteresis induced by friction in a single-degree-of-freedom system. The resulting system was modeled as a linear system with Duhem feedback. For each friction model, we computed the corresponding hysteresis map. Next, we developed a dc servo motor testbed and performed motion experiments. We then modeled the testbed dynamics and simulated the system using all three friction models. By comparing the simulated and experimental results, it was found that the LuGre model provides the best model of the g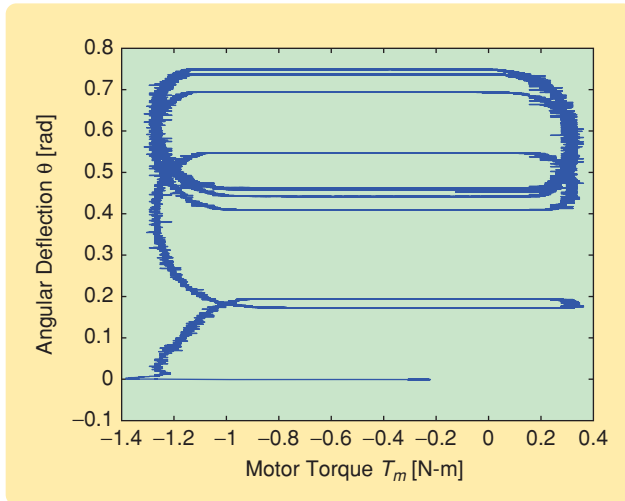earbox friction characteristics. A manual tuning approach was used to determine parameters that model the friction in the dc motor.

ACKNOWLEDGMENTS

REFERENCES

[1] B. Armstrong-Hélouvry, *Control of Machines with Friction*. Boston, MA: Kluwer, 1991.

[2] B. Armstrong-Hélouvry, P. Dupont, and C. Canudas de Wit, "A survey of model, analysis tools and compensation methods for the control of machines with friction," *Automatica*, vol. 30, no. 7, pp. 1083–1138, 1994.

[3] B. Feeny, A. Guran, N. Hinrichs, and K. Popp, "Historical review on dry friction and stick-slip phenomena," *Appl. Mech. Rev.*, vol. 51, no. 5, pp. 321–341, 1998.

[4] M.A. Krasnoselskii and A.V. Pokrovskii, *Systems with Hysteresis*. New York: Springer-Verlag, 1980.

[5] J.W. Macki, P. Nistri, and P. Zecca, "Mathematical models for hysteresis," *SIAM Rev.*, vol. 35, no. 1, pp. 94–123, 1993.

[6] P. Dahl, "Solid friction damping of mechanical vibrations," *AIAA J.*, vol. 14, no. 2, pp. 1675–1682, 1976.

[7] D.D. Rizos and S.D. Fassois, "Presliding friction identification based upon the Maxwell slip model structure," *Chaos*, vol. 14, no. 2, pp. 431–445, 2004.

[8] F. Al-Bender, V. Lampaert, and J. Swevers, "The generalized Maxwell-slip model: A novel model for friction simulation and compensation," *IEEE Trans. Autom. Contr.*, vol. 50, no. 11, pp. 1883–1887, 2005.

[9] V. Lampaert, J. Swevers, and F. Al-Bender, "Modification of the Leuven integrated friction model structure," *IEEE Trans. Autom. Contr.*, vol. 47, no. 4, pp. 683–687, 2002.

[10] C. Canudas de Wit, H. Olsson, K.J. Åstrom, and P. Lischinsky, "A new model for control of systems with friction," *IEEE Trans. Autom. Contr.*, vol. 40, no. 3, pp. 419–425, 1995.

[11] J. Swevers, F. Al-Bender, C.G. Ganseman, and T. Prajogo, "An integrated friction model structure with improved presliding behavior for accurate friction compensation," *IEEE Trans. Autom. Contr.*, vol. 45, no. 4, pp. 675–686, 2000.

[12] F. Al-Bender, V. Lampaert, and J. Swevers, "Modeling of dry sliding friction dynamics: From heuristic models to physically motivated models and back," *Chaos*, vol. 14, no. 2, pp. 446–445, 2004.

[13] F. Al-Bender, V. Lampaert, S.D. Fassois, D.D. Rizos, K. Worden, D. Engster, A. Hornstein, and U. Parlitz, "Measurement and identification of pre-sliding friction dynamics," in *Nonlinear Dynamics of Production Systems,* G. Radons and R. Neugebauer, Eds. Weinheim: Wiley, 2004, pp. 349–367.

[14] G. Ferretti, G. Magnani, and P. Rocco, "Single and multistate integral friction models," *IEEE Trans. Autom. Contr.*, vol. 49, no. 12, pp. 2292–2297, 2004.

[15] J. Amin, B. Friedland, and A. Harnoy, "Implementation of a friction estimation and compensation technique," *IEEE Contr. Sys. Mag.*, vol. 17, no. 4, pp. 71–76, 1997.

[16] S.C. Southward, C.J. Radcliffe, and C.R. MacCluer, "Robust nonlinear stick-slip friction compensation," *J. Dynam. Syst. Meas. Contr.*, vol. 113, no. 4, pp. 639–645, 1991.

[17] S.-W. Lee and J.-H. Kim, "Robust adaptive stick-slip friction compensation," *IEEE Trans. Ind. Electron.*, vol. 42, no. 5, pp. 474–479, 1995.

[18] R.-H. Wu and P.-C. Tung, "Studies of stick-slip friction, presliding displacement, and hunting," *J. Dynam. Syst., Meas. Contr.*, vol. 124, no. 1, pp. 111–117, 2002.

[19] M. Marques, *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*. Cambridge, MA: Birkhaüser, 1993.

[20] S.P. Bhat and D.S. Bernstein, "Finite-time stability of continuous autonomous systems," *SIAM J. Contr. Optimiz.*, vol. 38, pp. 751–766, 2000.

[21] S.P. Bhat and D.S. Bernstein, "Continuous finite-time stabilization of the translational and rotational double integrators," *IEEE Trans. Autom. Contr.*, vol. 43, no. 5, pp. 678–682, 1998.

[22] J. Oh and D.S. Bernstein, "Step convergence analysis of nonlinear feedback hysteresis models," in *Proc. American Control Conf.*, Portland, OR, 2005, pp. 697–702.

[23] D. Angeli and E.D. Sontag, "Multi-stability in monotone input/output systems," *Sys. Contr. Lett.*, vol. 51, pp. 185–202, 2004.

[24] D. Angeli, J.E. Ferrell, and E.D. Sontag, "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive feedback systems," in *Proc. Nat. Academy Science*, 2004, vol. 101, no. 7, pp. 1822–1827.

[25] J. Oh and D.S. Bernstein, "Semilinear Duhem model for rate-independent and rate-dependent hysteresis," *IEEE Trans. Autom. Contr.*, vol. 50, no. 5, pp. 631–645, 2005.

[26] A. Visintin, *Differential Models of Hysteresis*. New York: Springer-Verlag, 1994.

[27] P.A. Bliman, "Mathematical study of the Dahl's friction model," *Euro. J. Mech. Solids*, vol. 11, no. 6, pp. 835–848, 1992.

[28] Y.Q. Ni, Z.G. Ying, J.M. Ko, and W.Q. Zhu, "Random response of integrable Duhem hysteretic systems under non-white excitation," *Int. J. Non-Linear Mech.*, vol. 37, no. 8, pp. 1407–1419, 2002.

[29] N. Barabanov and R. Ortega, "Necessary and sufficient conditions for passivity of the LuGre friction model," *IEEE Trans. Autom. Contr.*, vol. 45, no. 4, pp. 830–832, 2000.

[30] J. Swevers, F. Al-Bender, C. Ganseman, and T. Projogo, "An integrated friction model structure with improved presliding behavior for accurate friction compensation," *IEEE Trans. Autom. Contr.*, vol. 45, no. 4, pp. 675–686, 2000.

[31] F. Altpeter, "Friction modeling, identification and compensation," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 1999 [Online]. Available:http://biblion.epfl.ch/EPFL/theses/1999/1988/EPFL_TH1988.pdf.

[32] S.L. Lacy, D.S. Bernstein, and S.P. Bhat, "Hysteretic systems and step-convergent semistability," in *Proc. American Control Conf.*, Chicago, IL, June 2000, pp. 4139–4143.

AUTHOR INFORMATION

Ashwani K. Padthe received the bachelor's degree in aerospace engineering from the Indian Institute of Technology Bombay in 2003. In 2006 he received the M.S. in aerospace engineering at the University of Michigan, where he is now pursuing the Ph.D. His research interests include hysteretic systems and aeroelasticity.

Bojana Drincic received the bachelor's degree in aerospace engineering from the University of Texas at Austin in 2007. She is now pursuing a Ph.D. at the University of Michigan. Her research interests include hysteretic systems, systems with friction, and spacecraft dynamics and control.

JinHyoung Oh received the bachelor's degree in control and instrumentation engineering from Korea University and master's degrees in aerospace engineering and applied mathematics from Georgia Institute of Technology. He received the Ph.D. in aerospace engineering from the University of Michigan in 2005. He is currently employed by AutoLiv as a research engineer.

Demosthenis D. Rizos received the diploma in mechanical engineering from the University of Patras in 1999, where he is a doctoral candidate. His research interests are in experimental modal analysis, identification of linear and nonlinear mechanical systems, systems with friction, and fault detection and identification. He has contributed to several sponsored research projects and is a coauthor of more than 15 publications.

Spilios D. Fassois received the diploma in mechanical engineering from the National Technical University of Athens, Greece, in 1982, and the Ph.D. in mechanical engineering from the University of Wisconsin-Madison in 1986. He served on the faculty of the Department of Mechanical Engineering and Applied Mechanics of the University of Michigan from 1986 to 1994 and is currently on the faculty of the Department of Mechanical and Aeronautical Engineering of the University of Patras, Greece. His research interests are in the area of stochastic mechanical systems, including vibrating systems, with an emphasis on identification and fault detection. He leads the Stochastic Mechanical Systems and Automation (SMSA) Laboratory at the University of Patras and is the author of more than 130 publications. His research has been supported by Ford, Eastman Kodak, General Motors, Whirlpool, Hellenic Aerospace, Hellenic Railways, Alenia, BAE Systems, and Volkswagenstiftung. He is a member of the editorial board of the *Journal of Mechanical Systems and Signal Processing*, and he was a guest editor for the special section on system identification in the October 2007 issue of *IEEE Control Systems Magazine*.

Dennis S. Bernstein (dsbaero@umich.edu) is a professor in the Aerospace Engineering Department at the University of Michigan. He is editor-in-chief of *IEEE Control Systems Magazine* and the author of *Matrix Mathematics* (Princeton University Press). His interests are in system identification and adaptive control for aerospace applications. He can be contacted at the University of Michigan, Aerospace Engineering Department, 1320 Beal Ave., Ann Arbor, MI 48109-2140 USA.

Spreadsheet-Based Control System Analysis and Design

A Command-Oriented Toolbox for Education

NOURDINE ALIANE

Computer simulations are widely used in control education and constitute a natural complement to theoretical material. This statement is particularly true for an introductory course in control, which exposes students to abstract concepts that are difficult to grasp quickly [1]. In these circumstances, computer simulations can provide insight into the fundamental concepts and offer a valuable instrument to facilitate learning.

In control education, numerous professional software packages such as Matlab/Simulink and LabView have been adopted as teaching tools. However, to enhance learning, many educators have developed additional educational tools [2]–[7] that provide intuitive and interactive user interfaces. Interactivity, which means that users can dynamically change the system parameters and immediately see their effect on the system behavior, is a useful feature in control education [1]. Such interactivity can be found in educational tools such as *CCSdemo* [2], *Ictools* [3], *LSAD* [4], as well as in the more recent suite *PID-Basics, PID-Loop-Shaping*, and *PID-Windup* [5].

Developing simulation tools requires specialized software, such as Matlab or Sysquake [8], which is a Matlab dialect with interactive support. In addition to implementing the simulator engine (the solver), the development of such tools involves constructing a graphical user interface (GUI), which requires knowledge of advanced programming techniques such as event handling or multitasking. The creation of GUIs using only Matlab requires programming skills

and effort. Furthermore, adding a basic level of interactivity to Matlab applications is a time-consuming activity.

In this article, we use Microsoft Excel as an alternative platform for developing interactive learning tools for control education. Modern spreadsheets include a macro language that permits the inclusion of standard computer code and allows users to incorporate specific programs to manage worksheets. In the case of Microsoft Excel, its built-in Visual Basic for Application (VBA) has the flexibility of a general-purpose programming language, thereby enabling developers to greatly extend its capabilities by designing specific add-ins [9].

Spreadsheets can also be used as teaching tools for constructing computer demonstrations and laboratory simulations in many areas of engineering education [10]–[14]. The combination of their flexible programmable environment and plotting capabilities makes them powerful didactic tools. Spreadsheets' ubiquity and low cost compared to professional software permit us to consider them as an alternative tool for education.

In control education, however, the use of spreadsheets has received little attention. Some work can be found where spreadsheets are used as an alternative tool for simulating linear systems [15], digital control systems [16], and nonlinear systems [17]. In fact, simple and complex simulations of continuous, discrete, and hybrid systems can be easily performed.

In this article we present a spreadsheet-based tool for control education. This tool is a command-oriented control toolbox (COCT) for working with linear systems in the Excel environ-

ment. Currently, COCT supports single-input/single-output linear systems described by transfer functions and provides many of the essential utilities necessary for an introductory course in control. The COCT workbook is freely available on the Web [18].

WHAT IS COCT?

COCT is an Excel workbook based on a set of macros and functions written in the VBA language. These functions are organized as a toolbox that provides utilities for handling control objects such as transfer functions, time and frequency responses, closed-loop simulations, and other minor objects. Once activated, these functions are integrated within Excel, behaving like any built-in function. The COCT workbook is embedded in a single Excel file, and its use does not require installation or configuration of any kind.

In the COCT environment, the cells of worksheets are used as containers for entering and storing control objects. Consequently, the active worksheet acts as a user interface for control system analysis and design, negating the need to design a specific input form. Control objects are entered into cells in the same way as data are entered into spreadsheets. Additionally, several models and simulations can be entered into the same worksheet, and the user can switch from one simulation to another by moving the focus from one cell to another and running the corresponding simulations. Figure 1 shows a screenshot of a COCT worksheet with multiple models and simulations. To manage the COCT environment and make it user friendly, several mechanisms, such as

initializing, running, or focusing simulations, are programmed as VBA macros and arranged as a floating toolbar, which is automatically displayed when the COCT workbook is opened (see Figure 1).

Motivation

The development of COCT is motivated by the need to provide teachers and students with a learning tool that enables them to focus exclusively on control system concepts rather than on programming tasks. COCT is intended to be interactive in the sense that it allows users to see the effects of varying parameter values on the system behavior. Interactivity stimulates the students to learn and is therefore a key to enhancing their intuition about control fundamentals. Furthermore, COCT is designed to speed up learning by reducing the time needed to design and run simulations. Finally, this tool is intended to complement rather than replace existing approaches, such as lectures and practical laboratory assignments.

SCOPE OF COCT

COCT provides utilities that cover the fundamental control concepts for continuous-time linear systems. Users can enter transfer functions into a worksheet, plot the time and frequency responses to understand how the systems behave, explore how the loca-
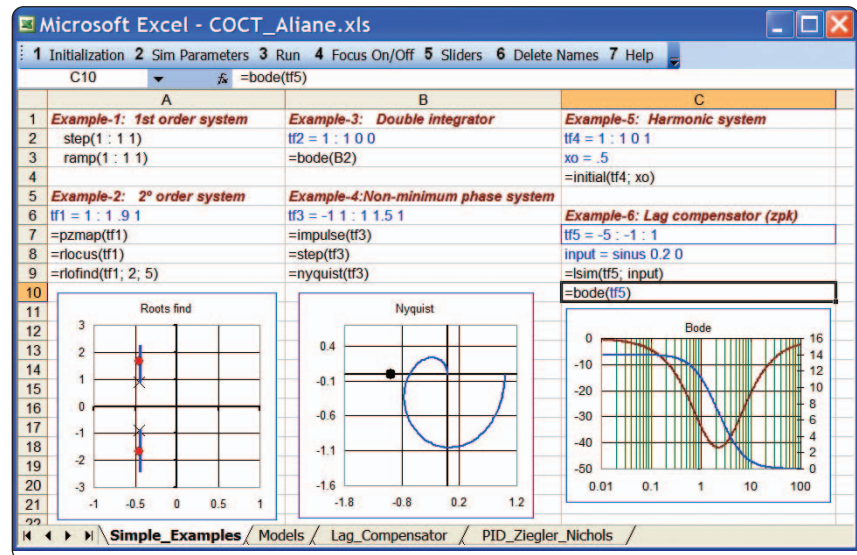

FIGURE 1 The command-oriented control toolbox (COCT) environment. The worksheet is configured to display the formulas in Excel standard format. The worksheet contains six examples. In the first column, the step response of the first-order system $G(s) = 1/(s + 1)$ is entered in cell A2. The step response is displayed by making the cell A2 active and then running the simulation by either clicking on the Run button or by pressing the Enter key. Management of the COCT environment is done through functions arranged in a floating toolbar.

tions of poles and zeros in the complex plane affect the time and frequency responses, or observe the effects of adding poles and zeros to systems, to mention only a few examples. COCT also provides utilities for simulating feedback control systems. The user can specify a model, tune the controller parameters interactively, and verify the performance in terms of rise time, steady-state error, or gain and phase margins. The principal capabili-

ties of COCT are listed in Table 1. Most of these capabilities are handled through model parameterization and interactivity, as discussed below.

Model Parameterization

When using COCT, transfer functions can be defined in rational polynomial or zero-pole-gain forms and can be entered into a worksheet in multiple ways. For example, a simple way to enter a transfer function is by typing it

---

Table 1 Scope of command-oriented control toolbox (COCT). COCT supports many commonly used functions for control system analysis and design. The user can specify a model, plot its time or frequency response, simulate a feedback control system, and tune the controller parameters.

1) Model definition and conversion
   Polynomial and zero-pole-gain forms
   Model conversion
   Series connection

2) Time domain
   Step response
   Impulse response
   Ramp response
   Initial response
   Sinusoidal response

3) Root locus
   Pole-zero map

4) Frequency domain
   Bode plots
   Gain and phase margin calculation
   Nyquist diagrams

5) Closed loop
   Feedback using an arbitrary transfer function
   PID control
   Time response to input reference and load disturbances

6) Relevant features
   Maintaining multiple simulations
   Model parameterization
   Interactivity

in a cell as a text entry. Alternatively, transfer functions can be assigned names, which can be used as references rather than using cell addresses in formulas. The convention adopted in COCT is that any cell with an entry in the form "alias = item" is assigned "alias" as a name. Additionally, the naming operation is programmed as a macro so that transfer functions are assigned names when they are entered, instead of the traditional way, that is, Insert->Name->Create. Named transfer functions are particularly useful for creating well-organized worksheets. Cells with transfer functions can be referenced anywhere in the workbook by functions that perform simulations and control calculations.

COCT also allows for entering models into a worksheet in a parameterized way so that their parameters can be changed in a straightforward manner. This versatile approach in creating models is achieved by two custom functions "=tf()" and "=zpk()," which can accept input parameters defined in individual cells or a range of cells. An open and parameterized model is constructed by entering its parameters in separate cells. Consequently, any changes in these parameters update the model automatically. Furthermore, the parameters of a model can be obtained using Excel formulas. Figure 2 shows an example of three parameterized models.
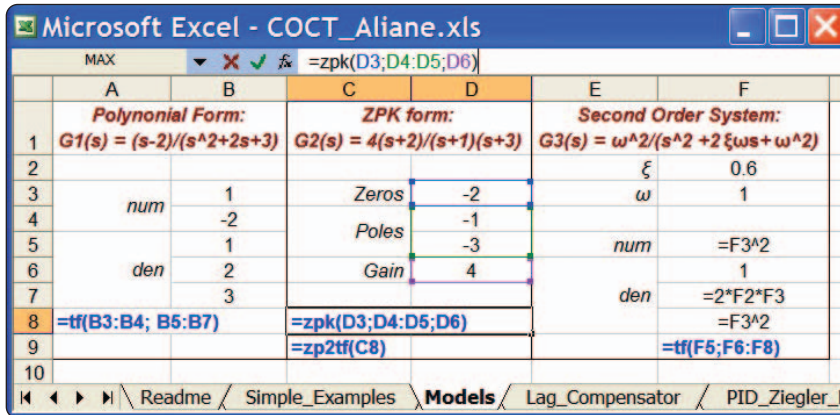
Essential Control System Functions

COCT provides several functions for analyzing linear systems in both the time and frequency domains. The major topics included are time responses to common input signals such as impulse, step, ramp, and sinusoid; frequency responses using the Bode and Nyquist plots; and additional utilities such as root locus or pole-zero maps. In the COCT environment, the time and frequency responses are identified as plot objects, which are obtained by using specific functions such as "=step()," "=bode()," "=nyquist()," or "=rlocus()." These functions are used much like any of the Excel built-in functions. A plot object is displayed by making its cell active and running the corresponding simulation.

COCT also provides utilities for simulating feedback control systems.

FIGURE 2 Model parameterization. Transfer functions $G1(s) = (s-2)/(s^2 + 2s + 3)$, $G2(s) = 4(s+2)/(s+1)(s+3)$, and $G3(s) = \omega^2/(s^2 + 2\xi\omega s + \omega^2)$ are entered as parameterized models in cells A8, C8, and F9, respectively. In the definition of $G3(s)$, the damping ratio ζ and the natural frequency ω are entered in separate cells, which are used in Excel formulas to define the model.
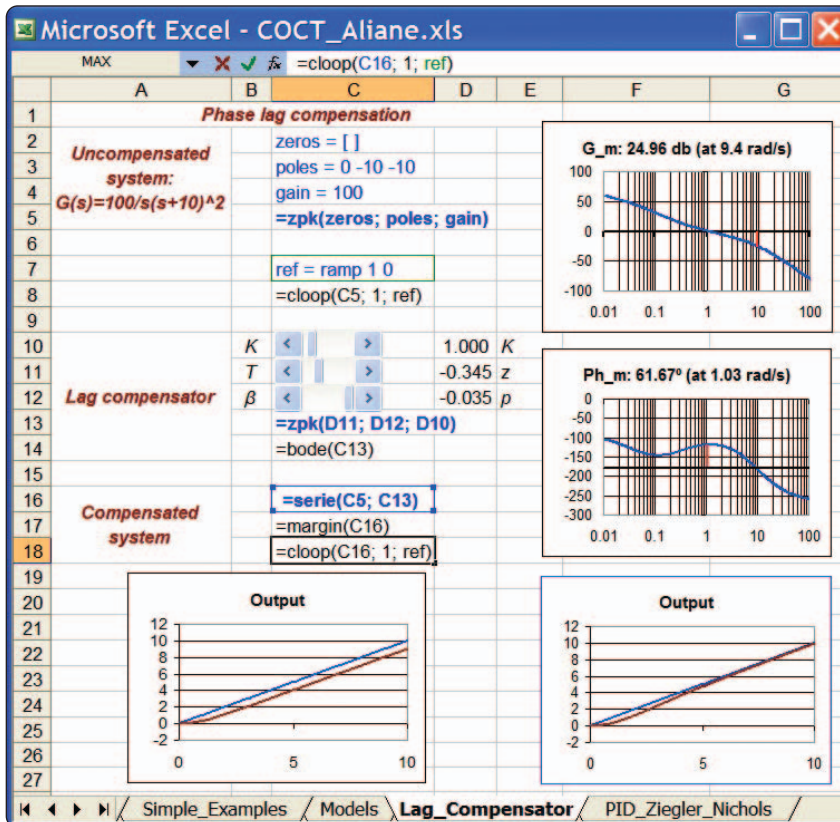

FIGURE 3 Phase-lag compensation. The transfer function of the system $G(s) = 100/s(s+10)^2$ is entered in cell C5. The lag compensator $C(s) = K(1 + Ts)/(1 + \beta Ts)$, with $\beta > 1$, is entered in C13. The series connection of the compensator and the system is entered in B16. The Bode plots with gain and phase margin calculation are entered in C17. The time responses to a ramp input of both uncompensated and compensated systems are entered in cells C8 and C18, respectively.

A closed-loop is defined by the "=cloop()" function, specifying the model, the controller, the input reference, and an optional external disturbance. PID controllers are of particular interest when studying control systems and constitute a required chapter in a basic control course. A PID controller is defined by the "=pid()" function specifying the three tuning parameters. The influence on the system behavior of each gain in the PID controller can be observed interactively by entering the tuning parameters in separate cells.

Furthermore, COCT provides a comprehensive tutorial on its use as well as a help section on how to use all of the supported functions.

Interactivity

The COCT environment supports interactivity in the sense that changes in any active element of a given simulation result in an immediate recalculation and presentation of the results. Thus, simulations with dynamic and animated graphs can be produced. This feature is achieved by focusing a cell containing a plot object, in which a programmed mechanism detects changes on the worksheet and subsequently runs the corresponding simulation.

To make the simulations even more interactive, elements such as sliders can be used to change the parameter values. This interface allows users to directly explore and understand the effect of varying parameters on the system behavior. For example, sliders can be used to investigate control system stability, analyze gain and phase margins, observe how the shape of both time and frequency responses are modified, adjust the controller's tuning parameters, and explore how the poles of a closed-loop system move along the root locus. In the COCT environment, the use of sliders does not require programming of any kind. The insertion of a slider into a worksheet is performed by selecting it from the (View->Toolbars->Control) menu and by dragging out an image. This feature makes the COCT environment
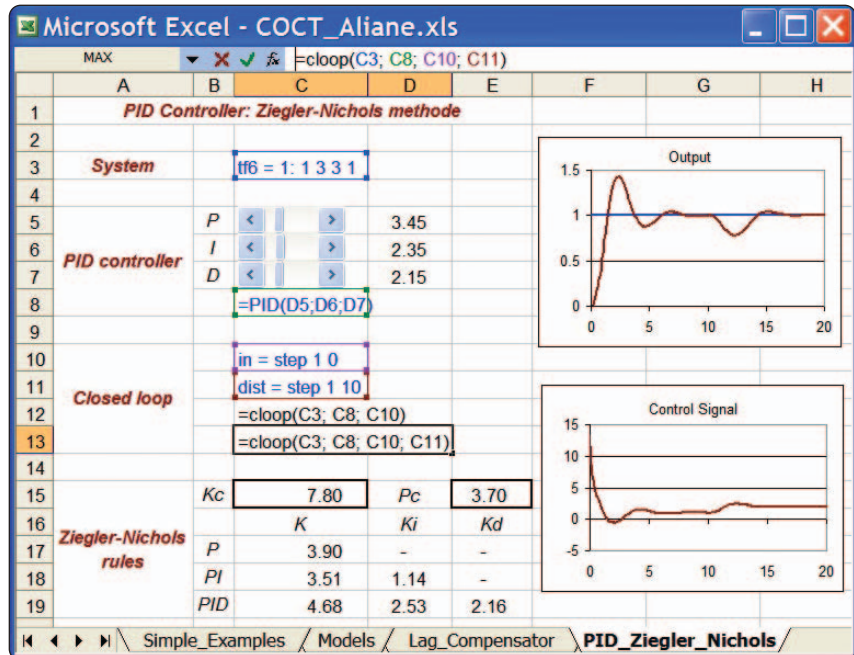

FIGURE 4 PID control. The transfer function of the system $G(s) = 1/(s+1)^3$ is entered in cell C3. The PID controller is entered in cell C8, and its tuning parameters are entered in separate cells. The simulation of the closed loop is entered in C13. The controlled system is subject to an input reference (a unit step applied at t = 0 s), and to an external disturbance (a unit step applied at t = 10 s), which are entered in cells C10 and C11, respectively. Ziegler-Nichols rules are also entered as standard Excel formulas.

highly interactive and enables users to perform "what-if" experiments in a straightforward manner.

EXAMPLES

This section presents two examples that illustrate some of COCT's capabilities. Although these examples do not fully exploit all of the COCT features, they illustrate how the iterative process involved in the design of classical controllers is carried out interactively.

Example 1: Phase-Lag Compensation

The purpose of phase-lag compensation in the frequency domain is to satisfy specifications on steady-state accuracy and phase margin. The design process is graphical in nature and consists of two steps. The first step is to determine the gain of the compensator to satisfy the steady-state error specification, while the second step is to determine the compensator's pole and zero locations that satisfy the phase-margin requirement.

COCT offers utilities to simplify the phase-lag compensation design process and help students understand its characteristics. One such simulation is shown in Figure 3. To visualize the design process, the compensator parameters are entered in separate cells and linked to sliders. Additionally, Bode diagrams showing the gain and phase margins as well as the time response to a ramp input of both uncompensated and compensated systems are visualized. Dragging the sliders immediately updates the plots, and, consequently, students can obtain an immediate sense of how moving a compensator pole or zero location influences the gain and phase margins.

Example 2: PID Controller and Ziegler-Nichols Method

In this example, we consider tuning a PID controller using the Ziegler-Nichols rules. According to the margin stability method, a PID controller is first tuned by setting it to the proportional mode and adjusting the gain

to make the system marginally stable. The final recommended values of the tuning parameters are obtained by finding the corresponding critical gain and the ultimate period of oscillation. This process is iterative and suitable for graphical treatment.

The Ziegler-Nichols method is illustrated graphically in the simulation shown in Figure 4. The PID parameters are entered in separate cells, and their values can be changed interactively. Ziegler-Nichols rules are also entered as Excel formulas. The design process is simplified by taking advantage of interactivity. The interaction is more straightforward due to the sliders, allowing students to clearly perceive changes in the performance. Furthermore, the user can improve the parameter tuning through the trial and error approach.

USAGE EXPERIENCE

In the 2006–2007 academic year, we projected COCT screenshots during lectures at Universidad Europea de Madrid. COCT capability of quickly generating multiple examples with plots of time and frequency responses proved to be helpful. Its interactive environment helps to draw comparisons between simulations, explore the effect of varying parameters, and give explanations through "what-if" scenarios. Many of the examples used in the lectures are collected in a single workbook. The fact that the COCT environment reduces the effort needed to specify and run simulations helps the instructor to proceed quickly. Finally, visibility of the entire worksheet highlights all of the parts of the simulations, and cells related to formulas are clearly indicated. This feature allows the instructor to quickly pose different cases in response to student questions.

The tool was also used by about 20 students as support material for exercises and as a basis for group discussions on more in-depth theoretical topics. Since the majority of students are familiar with Excel, they found COCT easy to learn and use. The stu-

dents agreed that its most valuable feature is its simplicity in setting sliders to explore the effects of varying parameters on the system behavior. In this aspect, sliders are used for interactively exploring control concepts. This sort of interaction encourages students to pay more attention and consequently helps them to gain better understanding of the qualitative aspects of control concepts.

FUTURE DEVELOPMENTS

Although the COCT toolbox is in its infancy, the current version marks the first step in developing a more complete control systems software package for the Excel environment. For example, the COCT kernel can be used to design specific demonstrations and interactive modules. However, the most challenging prospect is to explore many of the Microsoft Excel facilities for extending COCT's capabilities, for which we have two projects underway. The first one consists of interfacing COCT with external libraries such as LAPAK or EISPAK for performing large calculations, and the second one is related to interfacing COCT with external hardware for performing real-time control.

CONCLUSIONS

A spreadsheet-based tool for control education has been presented. The tool is a COCT for working with linear systems in the Excel environment. The tool provides the essential utilities for linear system analysis, controller design, and simulation of closed-loop systems. Currently, COCT supports only single-input, single-output linear systems described by scalar transfer functions.

COCT supports multiple simulations and system analysis on a single worksheet and allows users to switch from one simulation to another by moving the focus from one cell to another. Thanks to its interactivity, users can define "what-if" scenarios, in which they can dynamically change the system parameters and observe their influence on the system behavior.

Since COCT is embedded in a single Excel file, it does not require installation or configuration of any kind. The COCT environment reduces the effort needed to specify and run simulations. Moreover, experience with Excel is not required to develop helpful simulations with a good degree of interactivity.

REFERENCES

[1] S. Dormido, "Control learning: present and future," *Annu. Rev. Contr.*, vol. 28, no. 1, pp. 115–136, 2004.

[2] B. Wittenmark, H. Hägglund, and M. Johansson, "Dynamic pictures and interactive learning," *IEEE Contr. Syst. Mag.*, vol. 18, no. 3, pp. 26–32, 1998.

[3] M. Johansson, M. Gäfvert, and K. J. Åström, "Interactive tools for education in automatic control," *IEEE Contr. Syst. Mag.*, vol. 18, no. 3, pp. 33–40, 1998.

[4] S. Dormido, S. Dormido-Canto, R. Dormido-Canto, J. Sanchez, and N. Duro, "The role of interactivity in control learning," *Int. J. Eng. Edu.*, vol. 21, no. 6, pp. 1122–1133, 2005.

[5] J.L. Guzmán, K.J. Åström, S. Dormido, T. Hägglund, and Y. Piguet, "Interactive learning modules for PID control," in *Proc. 7th IFAC Symp. Advances in Control Education,* Madrid, Spain 2006 {CD-ROM].

[6] R.C. Garcia and B.S. Heck, "Enhancing classical controls education via interactive GUI design," *IEEE Contr. Syst. Mag.,* vol. 19, no. 3 , pp. 77–82, 1999.

[7] V. Kroumov, K. Shibayama, and A. Inoue, "Interactive learning tools for enhancing the education in control systems," in *Proc. 33rd ASEE/IEEE Frontiers in Education Conf.,* Boulder, CO, session T4E, 2003, pp. 23–27.

[8] Sysquake [Online]. Available: http://www. calerga.com/products/Sysquake/index.html

[9] D.M. Bourg, *Excel Scientific and Engineering Cookbook*, 1st ed. O'Reilly, 2006

[10] C. Rives and D.J. Lacks, "Teaching process control with a numerical approach based on spreadsheets," *Chem. Eng. Edu.,* vol. 36, no. 4. pp. 242–248, 2002.

[11] Chehab and A. El-Hajj, "Spreadsheet applications in electrical engineering: A review," *J. Comput. Applicat. Eng. Edu.,* vol. 20, no. 6, pp. 902–908, 2004.

[12] S.A. Oke, "Spreadsheet applications in engineering education: A review," *J. Comput. Applicat. Eng. Edu.,* vol. 20, no. 6, pp. 893–901, 2004.

[13] M. Schumack, "Use of a spreadsheet package to demonstrate fundamentals of computational fluid dynamics and heat transfer," *J. Comput. Applicat. Eng. Edu.,* vol. 20, no. 6, pp. 974–983, 2004.

[14] T.S. Hanania and I. Levin, "Spreadsheet-based logic controller for teaching fundamentals of requirements engineering," *Int. J. Eng. Edu.,* vol. 20, no. 6, pp. 939–948, 2004.

[15] A. El-Hajj, K.Y. Kabalan, and S. Khoury, "A linear control system simulation toolbox using spreadsheets," *IEEE Contr. Syst. Mag.,* vol. 20, no. 6, pp. 8–14, 2000.

[16] A. El-Hajj, S. Karaki, and K.Y. Kabalan, "Digital control systems simulation using

spreadsheets," *J. Comput. Applicat. Eng. Edu.*, vol. 11, no. 1, pp. 6–12, 2003.
[17] A. El-Hajj, S. Karaki, and Y.K. Karim, "A nonlinear control system simulation toolbox using spreadsheets," in *Proc. Int. Conf. Modelling and Simulation*, Pittsburgh, U.S.A, 2001, pp. 83–87.
[18] COCT workbook [Online]. Available: http://www.esp.uem.es/aliane/coct/coct.zip

AUTHOR INFORMATION
Nourdine Aliane (nourdine.aliane@uem.es) obtained the electrical engineering degree in 1990 from the École Nationale Polytechnique d'Alger and the Ph.D. in physics in 1995 from the Universidad Complutense de Madrid. He is currently a professor of control engineering and robotics at the Universidad Europea de Madrid, Spain. His interests include control systems, robotics, and education.

Advice to Young Researchers

MALCOLM D. SHUSTER

A wise man can learn from anyone, even from a fool, but a fool can learn from no one, not even from a wise man.
— Jewish proverb

It is easier to give advice than to bear one's own problems.
— Euripides (480–406 BCE)

A new researcher in any discipline is often at sea without a compass. He may have demonstrated tremendous skill already as a graduate student in the ability to achieve new results in research, but is unaware of the ethical requirements of research or of many activities that can improve the quality of his work and protect that work to a degree from serious errors. I am sorry to say that too many supposedly seasoned researchers are unaware of these errors or choose to ignore them. For the engineer there is the additional requirement that his research should have some connection to the real world, particularly on how one finds research problems in it. The present essay is not intended as a survivor's manual. It has loftier goals for those young creative researchers who, to steal a phrase from Faulkner [1], will not only survive but prevail, whether in industry, government, or academia. The present article is adapted from part of an earlier article [2].

Digital Object Identifier 10.1109/MCS.2008.929065

A BASIC RULE

Work hard and do good work! There is no substitute for this nor compensating factors. Also, to do really good work, you must be a little crazy. As the Romans said: *nullum magnum ingenium sine mixtura dementiae fuit* (there has not been great genius without an admixture of insanity), which need not imply that if you are a bit crazy, you will become a great genius, but it helps. But genius is not enough. My own career is certainly proof that persistence and hard work can compensate for lack of genius, unless one accepts Edison's definition of genius as one percent inspiration and ninety-nine percent perspiration, although my own ratio is certainly much poorer than Edison's.

RESEARCH

Be focused! It is important, certainly, that your research have breadth. If you are just starting out, breadth will get you better job offers, or present you with a greater range of areas in which to obtain research funding, but proven depth is more likely to get you fame. If you are a young academician, both will help you get tenure. It is important that there be at least one area where you have real depth, although it may take a long time to develop.

Put not thy trust in drawings! Drawings are very helpful in research. They excite our visual perception of the problem at hand, give us new insights, and are excellent ancillary devices for communication and teaching as well as mnemonics for important results. But drawings are also fraught with dangers, since it is often very difficult to interpret signs correctly from drawings. The misinterpretations can sometimes be very subtle. When my calculation almost "works" except for a sign inconsistency that I think I can fix later, that is usually the time for me to start doing the algebra microscopically. A word to the wise…

Put not thy trust in others! Never trust a published formula. Always rederive it yourself. Not only do published formulas sometimes contain errors (*mea culpa*), but we often understand the range of application of a formula only by deriving it ourselves and seeing then the hidden assumptions. If an article is not particularly well written, be prepared that there may be communication problems in the results as well. Always save your notes.

Be real! The best research often comes from real problems. I think that much academic research is, well, academic. If you are a young academician, form close ties with a local company or with a government facility. You may find that your best ideas, even general theoretical ideas, arise from their needs. (My most important research has often been research and development.) These contacts may also become a good source of research funding, summer salaries, and jobs for your students.

Don't always be practical! Sometimes impractical research can be very

spreadsheets," *J. Comput. Applicat. Eng. Edu.*, vol. 11, no. 1, pp. 6–12, 2003.
[17] A. El-Hajj, S. Karaki, and Y.K. Karim, "A nonlinear control system simulation toolbox using spreadsheets," in *Proc. Int. Conf. Modelling and Simulation*, Pittsburgh, U.S.A, 2001, pp. 83–87.
[18] COCT workbook [Online]. Available: http://www.esp.uem.es/aliane/coct/coct.zip

AUTHOR INFORMATION
Nourdine Aliane (nourdine.aliane@uem.es) obtained the electrical engineering degree in 1990 from the École Nationale Polytechnique d'Alger and the Ph.D. in physics in 1995 from the Universidad Complutense de Madrid. He is currently a professor of control engineering and robotics at the Universidad Europea de Madrid, Spain. His interests include control systems, robotics, and education.

Advice to Young Researchers

MALCOLM D. SHUSTER

A wise man can learn from anyone,
even from a fool, but a fool can
learn from no one, not even
from a wise man.
—Jewish proverb

It is easier to give advice than to
bear one's own problems.
—Euripides (480–406 BCE)

A new researcher in any discipline is often at sea without a compass. He may have demonstrated tremendous skill already as a graduate student in the ability to achieve new results in research, but is unaware of the ethical requirements of research or of many activities that can improve the quality of his work and protect that work to a degree from serious errors. I am sorry to say that too many supposedly seasoned researchers are unaware of these errors or choose to ignore them. For the engineer there is the additional requirement that his research should have some connection to the real world, particularly on how one finds research problems in it. The present essay is not intended as a survivor's manual. It has loftier goals for those young creative researchers who, to steal a phrase from Faulkner [1], will not only survive but prevail, whether in industry, government, or academia. The present article is adapted from part of an earlier article [2].

A BASIC RULE
Work hard and do good work! There is no substitute for this nor compensating factors. Also, to do really good work, you must be a little crazy. As the Romans said: *nullum magnum ingenium sine mixtura dementiae fuit* (there has not been great genius without an admixture of insanity), which need not imply that if you are a bit crazy, you will become a great genius, but it helps. But genius is not enough. My own career is certainly proof that persistence and hard work can compensate for lack of genius, unless one accepts Edison's definition of genius as one percent inspiration and ninety-nine percent perspiration, although my own ratio is certainly much poorer than Edison's.

RESEARCH
Be focused! It is important, certainly, that your research have breadth. If you are just starting out, breadth will get you better job offers, or present you with a greater range of areas in which to obtain research funding, but proven depth is more likely to get you fame. If you are a young academician, both will help you get tenure. It is important that there be at least one area where you have real depth, although it may take a long time to develop.

Put not thy trust in drawings! Drawings are very helpful in research. They excite our visual perception of the problem at hand, give us new insights, and are excellent ancillary devices for communication and teaching as well as mnemonics for important results. But drawings are also fraught with dangers, since it is often very difficult to interpret signs correctly from drawings. The misinterpretations can sometimes be very subtle. When my calculation almost "works" except for a sign inconsistency that I think I can fix later, that is usually the time for me to start doing the algebra microscopically. A word to the wise…

Put not thy trust in others! Never trust a published formula. Always rederive it yourself. Not only do published formulas sometimes contain errors (*mea culpa*), but we often understand the range of application of a formula only by deriving it ourselves and seeing then the hidden assumptions. If an article is not particularly well written, be prepared that there may be communication problems in the results as well. Always save your notes.

Be real! The best research often comes from real problems. I think that much academic research is, well, academic. If you are a young academician, form close ties with a local company or with a government facility. You may find that your best ideas, even general theoretical ideas, arise from their needs. (My most important research has often been research and development.) These contacts may also become a good source of research funding, summer salaries, and jobs for your students.

Don't always be practical! Sometimes impractical research can be very

enlightening about the foundations of one's field and even about the nature of real applications, but don't let impractical research become your dominant research theme.

A wise man can learn from anyone. Listen carefully to others, especially, to the questions of people struggling to understand your work. Sometimes their difficulties stem from situations not considered in your work or contain the germ of an important new idea. This has been the case for me in both Physics and Engineering. Keep in mind that the famous cameraman Gregg Toland volunteered his services on *Citizen Kane* to the untried director Orson Welles, because, he said, beginners had always provided him with the best new ideas.

It is better to be right than "practical." Wrong work is not "practical," just because you can describe it simply without equations. Rigorous mathematical work is not "only of theoretical interest," just because it requires a lot of equations. Do not let yourself be influenced by the kind of person who favors simplicity over correctness.

Pay attention to small details! Sometimes small items of no obvious consequence can lead to important research. Be especially watchful of steps that you must gloss over in a research paper, which you think are intuitively obvious, but which you don't know how to show. They may be a sign of hidden gold for future (or current) research. Also, they may indicate that your earlier intuitive assumptions weren't quite right.

Develop intuition! We avoid mistakes by having good intuition. We develop intuition by making mistakes.

Not all work is valuable. Just because a work is correct doesn't mean that it is valuable. It must be really useful or increase our understanding or pose a new problem as well. In a word, it must be truly worthwhile. Regrettably, valueless research sometimes (often?) gets published.

Be a dilettante! It is worthwhile to approach some research as a dilettante, that is, to do the work on your own, on your own schedule and not be tied to

contract or graduate-student timetables. Don't necessarily give your most original ideas to a Ph.D. student, unless you are certain that the student is at least as smart and as imaginative as you, because then the work will be done on the student's timetable and at the student's level of competence and creativity. To make the computer simulations a master's thesis after the theory has been worked out completely is another thing. Research cannot all be part of the *business* of being a professor. Some of it must be a truly joyous personal activity not easily given to a subordinate. If we forget this, then we risk making research just a job.

Be unreasonable sometimes! Being unrelentingly reasonable or politically correct in life (and in research) makes Jack a dull boy. Sometimes ethics forces us to go out on a limb, and our quest for truth forces us to explore territory that others would rather we avoid. Shakespeare's Polonius in *Hamlet* definitely had something to say here.

Knowledge is infinite; humans are finite. While it is good to study just to acquire knowledge, keep in mind that there is no limit to the amount of background one can acquire on a particular topic. Don't wait until you have complete knowledge of a topic before you begin to develop your own ideas. Usually, the idea comes first. We often learn what things we really need to learn *as* we do research. Sometimes it is even more efficient simply to "reinvent the wheel." Learning too much about a topic can make us unoriginal, because we will get stuck in the rut of previous work, even our own.

The most important research is often about finding questions, not about finding answers. As engineers or scientists, not only must we find the answers to important questions, we must find the important questions too. Computers and simulation can only be part of research. As Picasso once said, "Computers are useless. They can only give you answers." (Los ordenatores son inútiles. Sólo pueden darte respuestas.) That

statement is not altogether true for us, but there is much truth in it nonetheless.

Check your work! This doesn't mean only not finding errors when you reread your derivation or your computer program. Make certain that your new equations agree with trusted known results in special cases. In derivations, if you are able, derive your result in more than one way. In your computer output, check intermediate results. Check any properties your results should have. In simulation, test for simple models for which you know the answers or can easily calculate them by hand. Never just say: "I coded my equations, and this is what I got." In a batch estimation problem, for example, do your estimate errors decrease inversely as the square root of N, as N, the number of measurements, becomes large? Are two-thirds of the estimation errors smaller than the theoretical one sigma in magnitude?

Have courage! Do not be afraid to examine a topic, just because a respected colleague thinks such work is silly or that the problem has been settled. On the other hand, if you discover nothing important, move on. Most of my own early efforts in Astronautics did not lead to a publication.

Carpe diem! Seize the day, said the Roman poet Horace (Quintus Horatius Flaccus, 65–8 BCE). Do not procrastinate in your work. Above all, do not procrastinate when it comes to working independently. When one arrives at a faculty as a new assistant professor or even post-doctoral fellow, it is natural to want to be helpful to your new colleagues, or to become a secondary contributor to their work, which you may rightly regard as better than anything that you can do. Your rôle, however, is not to be effectively another graduate student for senior faculty, even if you have much to learn from them. Participate in their work, make yourself useful. It will not hurt your chances for tenure that some highly respected member of the faculty acknowledges a debt to you. But you must also give your department a real reason to give you tenure. You

must take the initiative in doing independent research. Don't procrastinate. You can find a million excuses to put this off. Remember the Nike commercial and *just do it!* If you are working in industry, and your boss asks you to do something, but you think you have a better idea, do first what he asks—you don't want to get a reputation of being uncooperative, which could get you fired or at least a poor raise—and then work out your own approach as well, but only if it can be done quickly. And don't show off afterward.

Keep it simple. Not all research is equal, and not all of it is valuable. The best research discovers simple things, which are often the hardest to find. It is easy simply to extend an existing piece of work to treat a small variation of a model or of a set of assumptions, or which consists mostly of simulations. Such work can be valuable, but the simple things are often the most important. Above all, avoid making a career out of publishing endless variations of your dissertation or some other research work. Your dissertation may contain unmined gold, but recognize when the vein has run out. Don't expend enormous effort just to develop a new methodology which is five percent better than an existing one. No one will care except you.

Be useful. Not every piece of work we do can be a breakthrough, but every piece of work should be useful. If not of immediate practical use, it should improve our understanding or provide an intermediate step to something useful and not simply demonstrate the brilliance of the researcher. Sometimes, we discover something new and end up changing a research area. But most research is not pathbreaking. Sometimes, a piece of research simply carries out a more basic and thorough analysis of previous work, not necessarily your own. Sometimes, one publishes afterthoughts about special cases. Sometimes, one publishes a survey or a tutorial. My most cited article isn't research at all but a survey paper,

which presented a great mass of material (two hundred years worth, in fact) from a unified point of view with consistent notation and conventions. When I submitted it, I told my friends that I had just submitted my most unoriginal paper, and which, I said, would become my most cited paper. I was right. The most important quality of good research is that it helps others rather than just satisfying one's own intellectual gratification.

Research ideas sometimes come from the strangest places. I have found that long-distance driving, walks in the mall, music, novels, poetry, and cooking sometimes stimulate research ideas. The list is endless. Read my earlier article on "The Arts and Engineering" in *IEEE Control Systems Magazine* [3].

SIMULATION IN RESEARCH

Simulation is a valuable tool. Simulation is valuable as a partial verification of your work, since simulation failure indicates that something is wrong somewhere. It is valuable also for illustrating your results and for determining the computational burden of an algorithm in real applications. In real-world applications, in which analytical verification often is not easily attainable in available time, simulation may be the best we can do to gain some (if not complete) confidence in a method. Keep in mind, however, that simulation experience is data, not insight or intuition, which come from physical or mathematical understanding. However, it is often the pathway to insight and intuition.

Thinking is better than computing. Often a simple analytical example is more illustrative and explains more than a numerical example.

Simulation is not a proof. I see too much shoddy work, sometimes even in the journals, "proved" by simulations. The worst sort of article, in my opinion, is the kind in which the writer proposes arbitrarily several different solutions to a problem, none of which is obviously correct mathematically, and decides which

one is best by simulation tests. Just because your residual errors converge to zero or some small value doesn't mean that your work is correct. The correct approach may converge faster or to a smaller value, or the asymptotic error level may be very different from what it should be as a result of errors in the approximation. Simulations to "verify" theoretical results should not just show that the errors become very small but that they have the anticipated values.

Not all simulations are equal. Sometimes I see illustrative simulation which simply repeats the steps of the author's *ad hoc* prescription, and illustrates only that the author has programmed the simulations correctly, although the model that was programmed may be wrong. Avoid this. Also, do not perform simulations which simply show the inner consistency of your work while avoiding numerical comparisons of your work with a known correct or more complete theory.

PUBLISHING

Write as you go! I have discovered that writing up my work may be my most important research tool. Generally, it is when I write that I discover the things which I should have done that I didn't consider doing originally or just didn't know how to do (but thought I did).

Don't rush to publication! I find that my publications in progress generally improve with age, provided I continue to revise them. If you are an assistant professor seeking tenure, this tactic may not work for you. All the same, walk, don't run.

Use clear and systematic notation! Do not introduce new notation just to be different. Using 0 for an index of an array is also inadvisable, since not every programming language permits a row or column index to have that value. Always enclosing the symbol for every rectangular matrix in brackets (once standard because of the limitations of typewriters) is also a bad idea, although it can add clarity in

special circumstances, as can any other delimiter. Likewise for underscoring the symbol for every column or row matrix. This occurs less frequently in the IEEE Controls arena than in other disciplines, such as Aerospace Engineering, where I work mostly.

Do not build permanent monuments to bad work! Conferences are a good place to present incomplete or not yet completely justifiable work; journals are not. (No place is a good place for work you know is incorrect.)

Don't defend your mistakes! If you have made a grievous error in a publication, especially in a journal article, don't try to cover up the mistake or, even worse, persist in it out of pride. A backlog of respect from previous good work may be squandered if you do, and you may be remembered more for your persistence in error than for more extensive good work. Better to publish an erratum or give notice of the error and correct it in a succeeding publication. I have done one or the other numerous times. No one will respect you less for having been honest.

The world will remember only your archival publications. With rare exceptions conference proceedings are eventually forgotten. Few people will go to the trouble of purchasing copies from the professional organization after the conference. If you wish your work to be remembered, publish it in a reputable journal. Note also that conferences *do not* review papers carefully. Don't betray your inexperience by boasting that your paper was presented at a "refereed" conference.

Good ideas often come quickly; good publications always require a lot of work. Some of my ideas come quickly, although I may spend a lot of time making the math come out right. (Nonetheless, my first significant researches in any new area were long, arduous, and frustrating.) Writing a journal article (and often a conference article) takes me forever. Good writing is actually part of the research. You don't understand a result until you can present it well.

Not quantity but quality! In Latin: *non multa sed multum*, or in Ancient Greek: οὐ πολλὰ αλλὰ πολὺ (literally, in both languages, "not many but much"). The ancients already had it right. Don't publish trivial or repetitive work or publish your work piecemeal just to get more publications. To do so in order to be able to attend a conference is excusable, but even there, when I see a chain of papers with few new results in each, bloated by unenlightening simulations, I am not impressed. At the same time, putting too many topics in one long paper can make it opaque. In that case, it can often be made clearer by dividing it into two (or more) shorter publications. I have tended to err on the side of articles that covered too many topics.

Be pedagogical in your papers! My papers most often have a strong tutorial element (meant largely for me), and I have often been accused of writing a textbook in the journals. That accusation may be justified, but I also get a lot of citations. It is easy to overdo pedagogy, and hard to find the right amount. Work at it. When you write a paper, you are not only reporting what you did but also teaching your readers how to do what you did. The journey, one might say, is part of the result. This may be too much to ask of young researchers, but it is worth a try.

Good cooks leave good recipes. Also helpful when presenting a very new method is to give a detailed bulleted prescription in a later section or in an appendix where the steps are summarized one by one. Don't repeat long equations, but simply refer to them by number in the main text. Except in rare instances, publishing code in Matlab or C is probably not a good idea, and a typographical error may make enemies for you down the road. Descriptive code is safer, and reading your paper should not be an experience similar to puzzling out a computer program.

Always give credit where credit is due! If you use someone else's results in your paper, always cite them fully and unambiguously, making clear what parts of your paper are taken from theirs. It is always better to err on the side of being overly scrupulous. Always check for earlier work before you publish a result. Given the bibliographic resources provided by the Internet, especially (in 2008) the AIAA, IEEE, Google, and Google Scholar Web sites, there is no excuse for being unaware of any important related work published in the last two decades. When I see an article that has no references more recent than 20 years ago, I become suspicious that the research is old and obsolete or that a superior similar article by someone else has already appeared.

Pride goeth before a fall. (approx. Proverbs [16:18]). Lack of pride goes nowhere. Worry not only about the value and correctness of your work. Be concerned also with its presentation, the writing, the drawings, the plots. Learn to write clearly. Every technical writer should read Strunk & White [4] once a year. Streamline and simplify the mathematics as much as possible. If you typeset your work, learn good typesetting style. *The Chicago Manual of Style* [5] is the common standard. Read extensively outside Engineering. The quality of expression from non-engineers is usually much better than ours. Remember, that when you write a paper, you are telling a story. A good paper, like a good short story or a good film, has a clear beginning, middle, and end. Writing a scientific paper is not like writing the great American novel, nor is it like ordering a pizza.

Non illigitimi carborundum est! Do not let yourself be overly angered by unfavorable reviews of your submitted articles. Most reviewers are careful and thoughtful. Occasionally, there is the mean-spirited review. If a reviewer has misunderstood your paper, you should examine the review carefully in order to decide whether he was simply unequipped to review your paper (this happens) or your presentation wasn't clear enough (this

happens more often). Innovative work is not always recognized right away. It is better to rewrite your article to make the nature or value of your innovation more apparent than to argue with the reviewer. The reviewer will probably not be the only reader to miss your point. Detailed reviews are worth gold, even if they are negative.

TEACHING

Good teaching in Engineering is Research. In the words of the philosopher Columbanus Sutor: *Docendo discimus*, by teaching we learn, and learning is certainly part of research. Preparing a course in your area of research is very much a research-like activity, if you do it right, and a source of ideas. This advice is almost a corollary of the statement that good writing is part of research. If you acquire a reputation as a great teacher, you will also make valuable friends among both students and faculty. One of my most valuable journal articles came from trying to motivate an algorithm for course presentation.

THE DARK SIDE OF RESEARCH

We all screw up. This is true of the great and the small. Some of my most admirable colleagues and even the author have violated many of these items of advice, mostly in our youth. It is my least admirable colleagues who continue to violate most of them, even in their mature years. The sky will not fall if you decide that you have made a mistake, even several mistakes. Life and work require constant adjustment.

Life is not fair. No one said that research would be easy. Don't give up too easily on a problem, and don't work on an unyielding problem for too long without doing other things as well. Be prepared that you may not always be rewarded as you deserve, and sometimes colleagues resentful of your abilities and accomplishments may consciously try to do you damage or try to take credit for your work. Research is carried out by people and

is subject, therefore, to the inequities of any human endeavor. To work hard and to do good work is often the best we can do. The most important praise we receive comes from within (and stays there). If we constantly produce work of high quality and have respect for ourselves, then eventually others will come to respect us. There is no other way.

Don't let the blues get you down! It is a truism that the most creative people are often the most susceptible to depression or even manic-depression. This is more frequent in the Arts than in Science, Engineering, and Mathematics, but it happens to us all the same. If you become depressed during a dry period, occupy yourself with other tasks, such as some less exciting work that has been on your shelf for some time, writing up unpublished work, studying a topic for which you had not previously found the time, writing a review of some research area (not necessarily your own and, perhaps, never to be published), developing software, or preparing a new course. All activities which lead to a desirable result stimulate us and create the endorphins which will take us out of the blues. Freud always contended that work is the best therapy.

LAST WORDS

Research isn't everything. When we are engaged in research, especially when we are working on our dissertations, we think that research is everything, but it is not. There is joy in discovering a new result, but I think research (in Engineering especially) is most satisfying when it serves some immediate practical purpose as well. I have done a lot of Engineering research, mostly in industry, all of it very satisfying; but, if I look back at more than thirty years in Engineering, my happiest were the first few, when research was not my objective nor part of my job description. Even for academics today, research is often only the icing on the cake, eaten in haste, and spoiled by deadlines, bureaucratic paperwork, and proposal writing.

Research has always been an important part of my life, but other aspects of my career, which were mostly development or management, have been just as rewarding if not more so.

When in doubt, do the right thing. We almost always know what the right thing is. Our moral dilemma generally is that we would rather, usually for selfish reasons, do something else. Morality requires courage and a willingness to give up something in order to do what is right. Loyalty, efficiency, and expediency are fine attributes, but they are not moral attributes. If we are to seek the truth, we must also be truthful ourselves in all things. Make the world a better place.

Take all advice with caution. All advice is based on the giver's personal experience and prejudices, the present advice no exception, and no advice can anticipate all situations. My advice includes practices that have worked for me and some others, and also warns against practices that I have found to lead to work which, in my opinion, is of diminished quality. Many of my close colleagues do not agree with every one of these items. Some things one must simply learn for oneself—the hard way. No writings can protect you from every disaster. If my counsels have made you think more about what you do, and especially if they have given you encouragement, then I am very pleased.

Above all, be happy in your work! Readers of my generation will recognize here the mantra of the Japanese commandant of the prisoner-of-war camp in the film *The Bridge on the River Kwai*, who hardly created a happy work environment. Research should be a source of joy, of exhilaration, and, in many ways, an act of love. If it isn't, then it may be difficult to endure the hardships that research entails. I abandoned a productive career in Nuclear Physics thirty years ago, largely because it stopped being fun. I never expected to do research in Astronautics—I was even looking forward to a break from

To be included in the conference calendar, send announcements to:

John Watkins
j.watkins@ieee.org

▲ Indicates CSS-sponsored conference
● Indicates CSS-cosponsored conference

The complete and current list of CSS-sponsored and cosponsored conferences is available at the CSS Web site, http://www.ieeecss.org.

» 2008

UKACC CONTROL CONFERENCE (CONTROL 2008)
2–4 September, Manchester, U.K.
General Chair: Hong Wang
Program Chairs: John O. Gray, Guoping Liu
http://www.control2008.org/index.php

▲ **MULTICONFERENCE ON SYSTEMS AND CONTROL**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzales
Program Chairs: Gary Balas (CCA), Marco Lovera (CACSD), Kevin L. Moore (ISIC)
http://conferenze.dei.polimi.it/msc08/

Digital Object Identifier 10.1109/MCS.2008.927315

▲ **CONFERENCE ON CONTROL APPLICATIONS**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzalez
Program Chair: Gary Balas
http://conferenze.dei.polimi.it//msc08/cca/cca_index.htm

▲ **COMPUTER-AIDED CONTROL SYSTEMS DESIGN**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzalez
Program Chair: Marco Lovera
http://conferenze.dei.polimi.it//msc08/cacsd/cacsd_index.htm

▲ **INTERNATIONAL SYMPOSIUM ON INTELLIGENT CONTROL**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzalez
Program Chair: Kevin Moore
http://conferenze.dei.polimi.it//msc08/isic/isic_index.ht

● **INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION, AND SYSTEMS 2008 (ICCAS 2008)**
10–14 October, Seoul, Korea
General Chair: Sung Kwun Kim
Program Chair: Doo Yong Lee
http://2008.iccas.org/

2008 DYNAMIC SYSTEMS AND CONTROL CONFERENCE (DSCC 2008)
20–22 October, Ann Arbor, Michigan, USA
General Chair: Galip Ulsoy
Program Chair: Eduardo Misawa
http://www.dsc-conference.org/

▲ **47TH IEEE CONFERENCE ON DECISION AND CONTROL**
9–11 December, Cancun, Mexico
General Chair: Chaouki Abdallah
Program Chair: Thomas Parisini
http://control.disp.uniroma2.it/CDC08/

● **10TH INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION, ROBOTICS AND VISION (ICARV 2008)**
17–20 December, Hanoi, Vietnam
General Chair: Y.C. Soh
Program Chair: C. Wen
http://www.icarcv.org/2008/

» 2009

▲ **2009 AMERICAN CONTROL CONFERENCE**
10–12 June, St. Louis, Missouri, USA
General Chair: K. Hoo

» FOCUS ON EDUCATION *(continued from page 117)*

research—but, it seems, research was unavoidable; it's my nature. May research bring you these same joys.

ACKNOWLEDGMENT
SCRIPTOR HVIVS ARTICVLI CORDE IPSO DIONYSIO ELECTRO MVLTIBVS EMMENDATIONIBVS GRATIAS AGIT.

REFERENCES
[1] W.C. Faulkner, "Nobel speech," in *The Faulkner Reader*. New York: Random House, 1954; reissue: The Modern Library (a division of Random House) 1977.

[2] M.D. Shuster, "In my estimation" *J. Astronautical Sci.*, vol. 54, nos. 3–4, pp. 273–297, July–Dec. 2006.

[3] M.D. Shuster, "The arts and engineering," *IEEE Control Syst. Mag.*, vol. 28, pp. 96–98, Aug. 2008.

[4] W. Strunk, Jr., E.B. White, and R.R. Angell, *The Elements of Style*, 4th ed. New York and London: Longman, 2000.

[5] *The Chicago Manual of Style*, 15th ed. Chicago, IL:: Univ. of Chicago Press, 2003.

AUTHOR INFORMATION
Malcolm Shuster (mdshuster@ comcast.net.) is director of Research for Acme Space Company. After a decade of research in theoretical Nuclear Physics, he wanted to do something different and spent the next 30 years mostly in the aerospace industry, where he quickly found himself doing research again. More extensive author information can be found in a previous article in the *IEEE Control Systems Magazine* (August 2008). He can be contacted at Acme Space Company, 13017 Wisteria Drive, Box 328, Germantown, MD 20874; http//home.comcast.net/~mdshuster.

Interactive Learning Modules for PID Control

Using Interactive Graphics to Learn PID Control and Develop Intuition

JOSÉ LUIS GUZMAN, KARL JOHAN ÅSTRÖM, SEBASTIAN DORMIDO, TORE HÄGGLUND, MANUEL BERENGUEL, and YVES PIGUET

Interactive tools can be used to complement books and lectures [1]–[4]. This article describes three interactive learning modules that are designed to develop intuition as well as a working knowledge of proportional-integral-derivative (PID) control. These three modules comprise a package called interactive learning modules for PID (ILM-PID). By illustrating concepts such as tuning, robustness, loop shaping, and antiwindup, ILM-PID can be used for demonstrations, exercises, and self-study.

The main objective of the interactive modules is to explain basic concepts of PID control without considering implementation aspects. Although most PID controllers are implemented as sampled-data control systems, analysis and design are traditionally performed in continuous time assuming that the sampling rate for subsequent digital implementation is sufficiently fast. Implementation issues, such as aliasing, selection of the sampling time, signal prefiltering, influence of the discretization algorithms, and bumpless parameter changes, may be the aim of a future interactive modules focused on implementation aspects for PID control.

The modules of ILM-PID have menus for selecting process transfer functions and controller structures. In addition, parameters can be set, and results can be stored and loaded. A graphic display of time and frequency responses is a central part. The plots can be manipulated directly by dragging points and lines and by using sliders. Parameters that characterize performance and robustness are displayed. Each module has two icons called Instructions and Theory. Instructions provides access to a document that contains suggestions for exercises, while Theory provides access to relevant theory by means of the Internet. The modules are implemented in Sysquake [5], a Matlab-like language with fast execution and capabilities for interactive graphics.

The following sections describe three modules that illustrate closed-loop fundamentals (PID Basics), loop-shaping design (PID Loop Shaping), and integrator windup (PID Windup). Readers are encouraged to visit the Web site [6] to experience the interactive features of ILM-PID. The modules are available for Windows, Mac,

and Linux operating systems and can be freely downloaded from the Sysquake Web site [7] as described in "Downloading and Using ILM-PID."

PID BASICS

The module PID Basics is designed to explore the properties of a simple feedback loop by showing the time and frequency responses of a closed-loop system and demonstrating how these responses are influenced by the choice of controller parameters.

A block diagram of a basic feedback loop is shown in Figure 1, where P and C are the process and controller transfer functions, respectively, and F is the filter transfer function for the setpoint. The system has three inputs representing the setpoint y_{sp}, the load disturbance d, and the measurement noise n. It is assumed that the load disturbance acts at the process input and that the measurement noise acts at the process output. The controller must reduce the effect of the load disturbance and make the process variable x follow the setpoint y_{sp}, while not injecting too much measurement noise. In addition, the closed-loop system must be insensitive to variations in the process dynamics.

At least three signals are of interest, namely, the process output signal x, the measured output signal y, and the control signal u. Tracing signals in the block diagram in Figure 1 gives the relations

$$
\begin{aligned}
X &= \frac{PCF}{1+PC}Y_{sp} + \frac{P}{1+PC}D - \frac{PC}{1+PC}N \\
&= FTY_{sp} + PSD - TN,
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
Y &= \frac{PCF}{1+PC}Y_{sp} + \frac{P}{1+PC}D + \frac{1}{1+PC}N \\
&= PCFY_{sp} + PSD + SN,
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
U &= \frac{CF}{1+PC}Y_{sp} - \frac{PC}{1+PC}D - \frac{C}{1+PC}N \\
&= CFSY_{sp} - TD - CSN,
\end{aligned} \tag{3}
$$

where capital letters denote Laplace transforms of the corresponding time functions, $S = 1/(1+PC)$ is the sensitivity function, and $T = PC/(1+PC)$ is the complementary sensitivity function. Notice that the input-output relations are completely characterized by the six distinct transfer

Downloading and Using ILM-PID

Interactivity, which is the main feature of the tools described in this work, is difficult to explain in written text. The best way to appreciate the tools is to use them. We strongly recommend that the reader download and use them in parallel with reading this article. Executable versions for PC, Mac, and Linux are freely available for download from the Calerga Web site [7]. No licenses are required, and the executable modules can be freely distributed to students and colleagues.
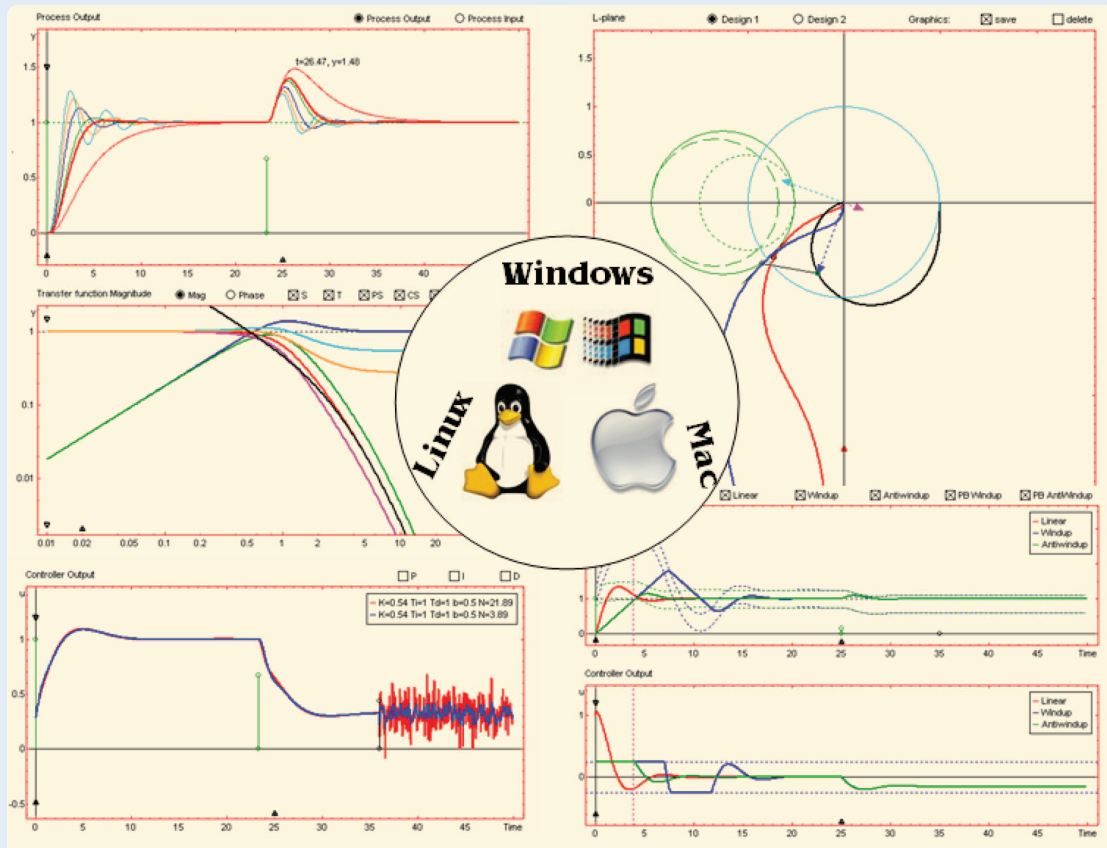
functions in (1)–(3). These transfer functions are called the gang of six in [8]. To analyze the closed-loop system it is necessary to consider all six transfer functions.

The time responses of the six transfer functions are illustrated by showing the response of the process output and control signals to a step in the setpoint, a step in the load disturbance, and wideband measurement noise, as illustrated in Figure 2. A mix of time and frequency responses can also be displayed.

Process models in the form of rational transfer functions with a time delay can be chosen from a menu that provides a collection of transfer functions. An arbitrary transfer function can also be entered using the standard Matlab format. The process gain and time delay can be changed interactively using sliders. The PID controller has the structure

$$U = K \left(bY_{sp} - Y + \frac{1}{sT_i}(Y_{sp} - Y) - \frac{sT_d}{1 + sT_d/N_d} Y \right),$$

where K is the proportional gain, T_i is the integral time, T_d is the derivative time, N_d is a parameter of the derivative term, and b is the setpoint weight.
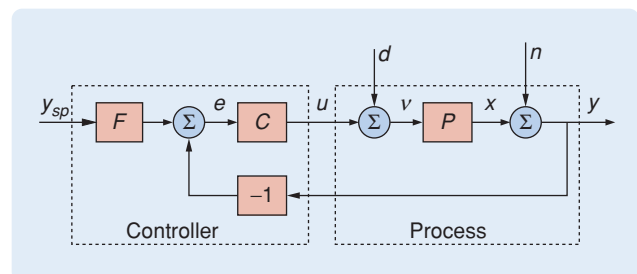
FIGURE 1 Basic feedback loop having two degrees of freedom. P and C are the process and controller transfer functions, respectively, and F is the filter transfer function on the setpoint. The variable y_{sp} is the setpoint, e is the tracking error, u is the controller output, d is the load disturbance, x is the process variable, n is the measurement noise, y is the measured output signal, and v is the controller output corrupted by the load disturbance d.

The Interactive Tool

The main screen of the tool is shown in Figure 3. The process is characterized by the parameter group located on the left-hand side of the screen, just below the icons (see Figure 3). The process is shown symbolically together with several interactive elements for changing the representative parameters of the process. The transfer function in Figure 3 is

$$G(s) = \frac{K_p}{(s+1)^n},$$

where the gain K_p and order n are the interactive elements, with numerical values $K_p = 1$ and $n = 4$.

When the user modifies any plant parameter, the symbolic representation of the process transfer function is immediately updated, and its effect is reflected on the remaining graphic elements.

Five buttons are available for selecting the desired controller. The buttons correspond to proportional (P), integral (I), proportional-integral (PI), proportional-derivative (PD), and proportional-integral-derivative (PID). Several sliders are available below the radio buttons for modifying the controller parameters. The number of sliders shown depends on the chosen controller. For instance, Figure 3 shows five sliders since the PID controller is selected.

Performance and Robustness Information

Parameters that characterize performance and robustness are also displayed on the screen. The performance criteria are based on the setpoint response, the load disturbance response, and the noise response. The setpoint response is characterized by the integral absolute error (IAE) and the

overshoot (overshoot). The load disturbance response is characterized by the integral absolute error (IAE), the integral gain $k_i = K/T_i$ (ki), the maximal error (emax), and the time to reach the maximum (tmax). The integral absolute errors and the maximal error values are normalized to unit step changes in setpoint and load disturbances. The response to measurement noise is characterized by the standard deviations of the process variable x (sigma_x), measured output y (sigma_y), and control signal u (sigma_u). The robustness measures are maximal sensitivity (Ms), maximal complementary sensitivity (Mt), gain margin (Gm), and phase margin (Pm). This information can be duplicated to compare two designs, as shown below. A more detailed description of these measures can be found in [8].

Graphics

Two graphics are shown on the right-hand side of the tool (Figure 3). Three representation modes can be selected from the Settings menu. These modes are time domain, frequency domain, and frequency/time domain.

The time domain mode is shown in Figure 3, where the time responses for the system output (Process Output) and input (Controller Output) are displayed. The initial part of the plots ($0 < t < 30$) shows the response to a step change in the setpoint represented by the transfer functions FT and CFS in (1)–(3). The middle portions of the plots ($30 < t < 60$) show the response to a step in the load disturbance represented by the transfer functions PS and T in (1)–(3). The last portions of the plots ($t > 60$) show the response to wideband measurement noise, which is represented by the transfer functions S and CS in (1)–(3).

Several elements on the graphics are available for interacting with the application. The vertical green line at time

FIGURE 2 Control system responses illustrating basic feedback system properties. To analyze the feedback loop, it is essential to consider six responses. These responses, which are referred to as the gang of six [8], are described by transfer functions in (1)–(3). One way to present this information is to show the process output y and the controller output u for step commands in setpoint and load disturbances, as well as the response to sensor noise, as shown here.
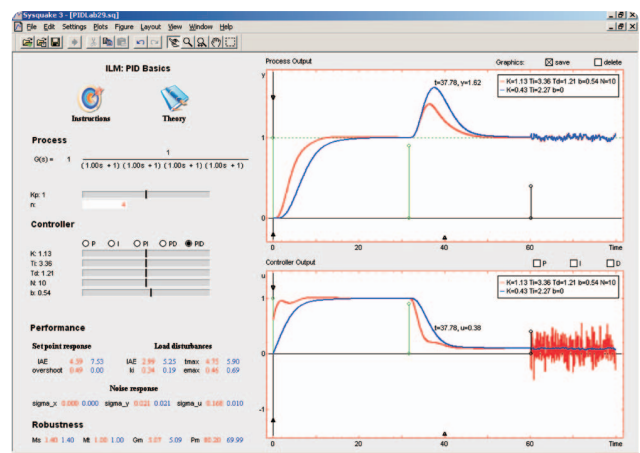

FIGURE 3 The user interface of the module PID Basics. The plots show the time response of the transfer functions in (1)–(3) [8]. Several graphical elements, shown on the same screen, are used to interactively analyze feedback fundamentals using PID control. This example provides a comparison between PI (blue) and PID (red) controllers.

$t = 0$ allows the setpoint amplitude to be modified. The green and black vertical lines located in the middle of the graphics allow setting the value and time instant for load disturbances and measurement noise, respectively. The vertical and horizontal scales can be changed using the black triangles (▲, ▼) available in the graphics. For instance, in Figure 3, the setpoint is set to one, the load disturbance is set to 0.9 at $t = 32$, and the measurement noise is set to 0.02 at $t = 60$. It is also possible to find the value for the input or output signal at a specific time by placing the mouse over the curve. Figure 3 shows an example in which, at the time instant $t = 37.78$, the output and input signals are 1.62 and 0.38, respectively. All of these options are available in both graphics, that is, Process Output and Controller Output.

The checkboxes save and delete above the Process Output graphic provide the ability to store a simulation for comparison. When the save button is selected, the current design is frozen and displayed in blue, and a new design in red appears, allowing the two designs to be compared. Performance and robustness parameters are duplicated, displaying the values in blue and red colors associated with each design. The Process Output and Controller Output graphics indicate the values of the controller parameters for both designs. Figure 3 presents an example that

compares the response of PI ($K = 0.43$, $T_i = 2.27$, $b = 0$) and PID ($K = 1.13$, $T_i = 3.36$, $T_d = 1.21$, $b = 0.54$, $N_d = 10$) controllers. Although the PID controller provides a better response to load disturbances by reacting faster, the noise also generates more control action. The delete option can be selected to remove a design. If the transfer function of the process or an input signal such as a setpoint, load disturbance, and measurement noise are altered, both sets of results are affected simultaneously. Only two designs are stored to keep the user interface simple.

Additional options for the time-domain mode are shown above the Controller Output graphic. These options show the proportional (P), integral (I), and derivative (D) signals of the controller.

The frequency domain mode is shown in Figure 4. When this mode is selected from the Settings menu, the left side of the tool remains unchanged. However, in this case the time responses are replaced by the magnitude and phase plots Transfer Function Magnitude and Transfer Function Phase. The vertical and horizontal scales can be interactively modified in the same way as in the time domain. The magnitude and phase for a specific frequency can be found by placing the mouse over the signals as shown in Figure 4(a).
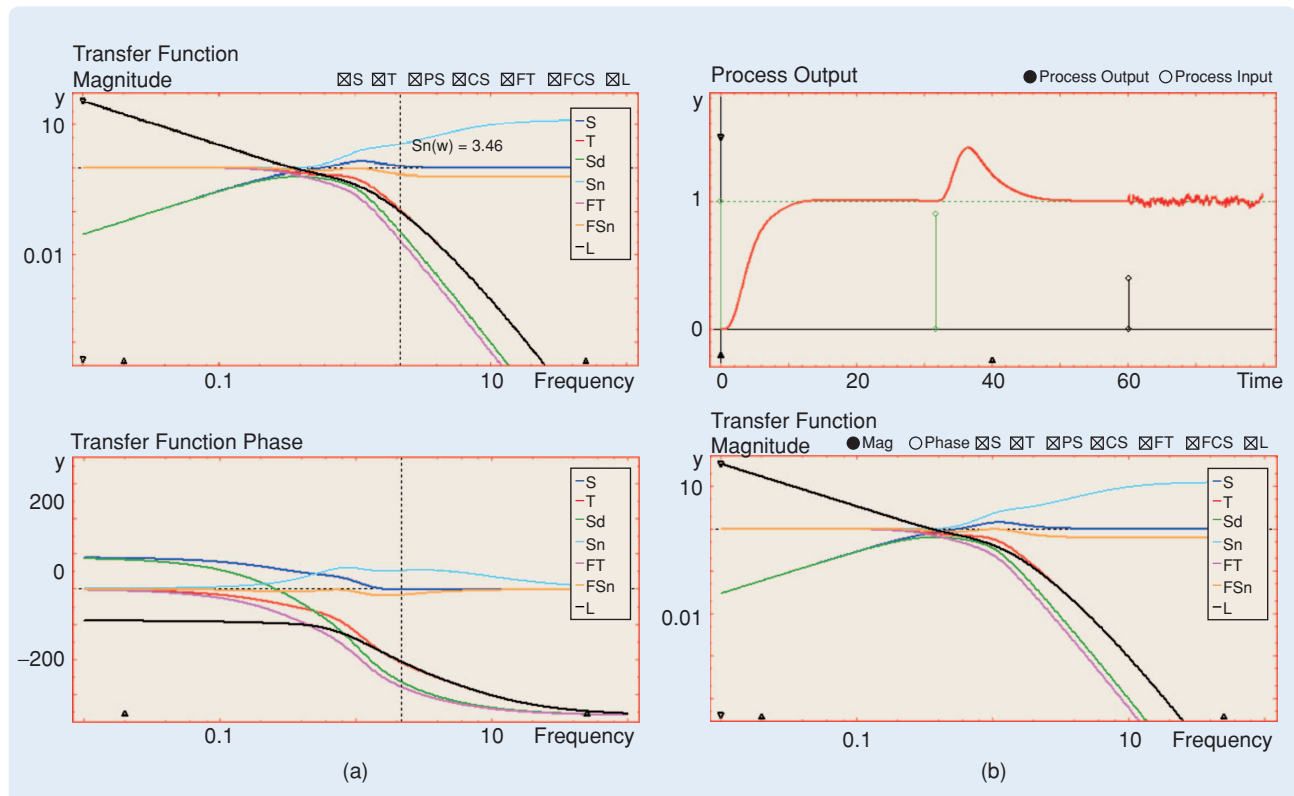

FIGURE 4 Time- and frequency-domain analysis using the interactive tool. (a) Frequency domain. The graphical part of PID Basics is shown for the frequency-domain mode, where the Transfer Function Magnitude and the Transfer Function Phase graphics are displayed. In this mode the user can study the transfer functions in (1)–(3) in the frequency domain using checkboxes placed above the Transfer Function Magnitude graphic. (b) Time and frequency responses, simultaneously. Above the graphics, the two buttons let the user choose between the output or input for the time domain, and magnitude or phase for the frequency domain.

The frequency response for the gang of six transfer functions and the open-loop transfer function $L(i\omega) = P(i\omega)C(i\omega)$ can be shown in the graphics using checkboxes placed above the Transfer Function Magnitude graphic. In Figure 4(a), all transfer functions are displayed.

Time and frequency responses can be shown simultaneously, as illustrated in Figure 4(b). The upper part represents the time responses, while the lower part shows the frequency responses. The default screen shows the output and the magnitude for the time and frequency domains, respectively. Above the graphics, the two buttons let the user choose between the output or input for the time domain and magnitude or phase for the frequency domain. This mode is useful since it is possible to view the effect of parameter modifications on both domains simultaneously.

Settings Menu

The Settings menu of PID Basics is divided into six groups. Arbitrary transfer functions can be selected using the first entry, Process Transfer Function. The numerator and denominator are introduced using a Matlab form.

Specific values for controller parameters can be entered using the Controller Parameters menu. Time and frequency responses can be selected from the third entry, Time/Frequency Domain, which has the options Time Domain, Frequency Domain, and Both Domains. The results can be stored and recalled using the Load/Save menu, which has the options Save Design and Load Design. All data on the screen can be saved using the option Save Report. From the menu selection Simulation the user can modify the simulation time, change the maximal time delay to avoid slow simulations, and activate the Sweep option to show the results for several controller parameters simultaneously. Parameters are swept between specified limits. This option is available only in the time-domain mode. When active, new radio buttons appear in the controller-parameters zone to permit the selection of the desired parameter to be swept. The last menu option, Examples Advanced PID Book, loads examples from [8], which the user can explore by modifying parameters.

Analysis and Control Design for Load Disturbances

Load disturbances are typically low-frequency signals that drive the system away from its desired behavior. The
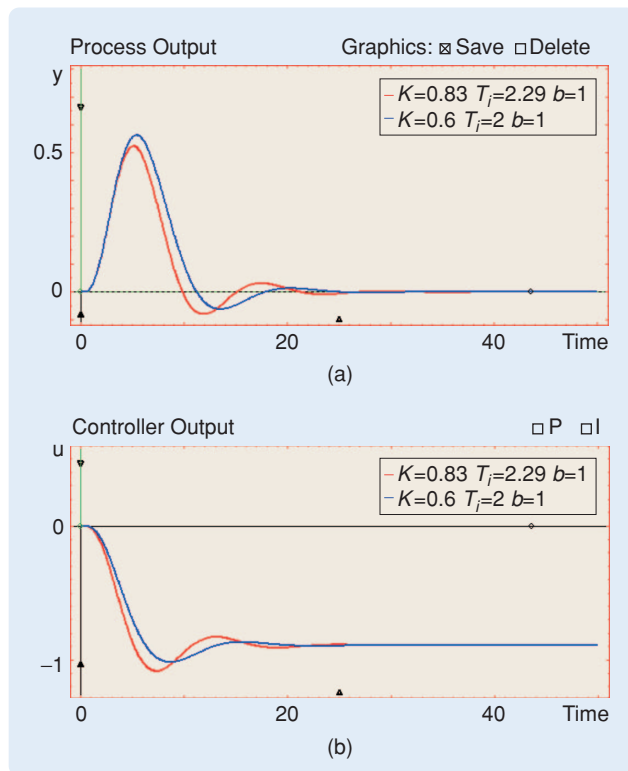
FIGURE 5 Load-disturbance response and influence of the integral gain k_i. For a system with $P(0) \neq 0$ and a controller with integral action, the low-frequency approximation is $G_{yd} \approx sP(0)/k_i$, where $k_i = K/T_i$ is the integral gain. For load disturbances with low-frequency content, the integral gain k_i is a measure of load-disturbance attenuation. The (a) process outputs and (b) control signals to load disturbances, respectively, are shown for two PI controllers with k_i values of 0.36 (in red) and 0.30 (in blue). The controller with larger integral gain provides a faster response to load disturbances.
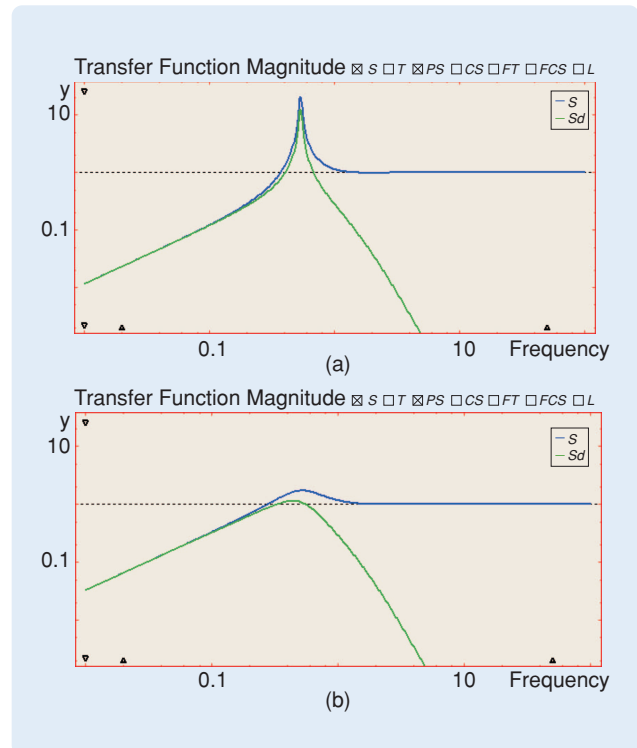

FIGURE 6 Frequency-domain interpretation of the load-disturbance response. Figure 5 shows that high values of the integral gain k_i provide better response to load disturbances. Although this rule is true, it must be used carefully. The frequency-domain responses of G_{yd} and S for two PI controllers with (a) $k_i = 0.85$ and (b) $k_i = 0.30$, respectively. As can be seen, large values of k_i imply large peaks of the sensitivity function $S = 1/(1 + PC)$. Therefore, a tradeoff occurs between load disturbance rejection and robustness.

response to load disturbances is a key issue in process control, since most controllers attempt to keep process variables close to desired setpoints [9]. The following example shows the effects of load disturbances and the influence of the controller parameters. The setpoint and noise amplitudes are set to zero, and the load disturbance is set to 0.9 at $t = 0$. The process transfer function is given by $G(s) = (s + 1)^{-4}$. The response of the process variable to load disturbances is given by the transfer function

$$G_{yd} = \frac{P}{1 + PC} = PS = \frac{T}{C}.$$

If $P(0) \neq 0$ and the controller has integral action, then the low-frequency approximation is $G_{yd} \approx sP(0)/k_i$, where $k_i = K/T_i$ is the integral gain. For load disturbances with low-frequency content, the integral gain k_i is a measure of load-disturbance attenuation. Figure 5 shows the load-disturbance responses for two PI controllers with k_i given by 0.36 (in red) and 0.30 (in blue). Although the controller with larger integral gain provides faster response and smaller values for IAE and emax to load disturbances, the stability margins are reduced. Figure 6 shows the frequency responses of G_{yd} and S for two PI controllers with large and small values of k_i (0.85 and 0.30, respectively). This figure reflects that large values of k_i imply large peaks of the sensitivity function. Therefore, a tradeoff becomes necessary between load-disturbance rejection and robustness.

Some tuning methods allow a tradeoff between robustness and load disturbance response. The approximate M-constrained integral-gain optimization (AMIGO) method [8], [10]–[12] maximizes integral gain under a robustness constraint; see "AMIGO Design Method." The result of applying AMIGO to this example is shown in Figure 7. The AMIGO-step method is used to design a PI controller with $K = 0.414$ and $T_i = 2.66$. The response to load disturbances is slower than the results presented in Figure 5, but stability margins result are improved, with $M_s = 1.32$ and $M_t = 1$.

PID LOOP SHAPING

This section briefly describes the main aspects of PID Loop Shaping. The main screen of the tool is shown in Figure 8.

Process

The process transfer function can be selected and modified depending on the option selected from the Settings menu.

AMIGO Design Method

Although load disturbances are often the major consideration in process control, robustness and measurement noise must also be considered. Requirements on setpoint response can be dealt with separately by using a controller with two degrees of freedom. The Ziegler-Nichols rules for tuning PID controllers are especially influential. These rules, however, have severe drawbacks, since they use insufficient process information and can yield closed-loop systems with poor robustness [11]. Loop shaping [13] can also be used for PID control, which gives a flexible design method that allows a tradeoff between performance and robustness. The design approach maximizes the integral gain subject to constraints on the maximum sensitivity. This method is called M-constrained integral gain optimization (MIGO) [8], [11].

AMIGO (approximate MIGO) design, which is a tuning method in the spirit of Ziegler and Nichols, is the result of finding simple tuning rules for the MIGO method. A large batch of representative processes is selected, including a wide variety of systems with essentially monotone step responses that are typically encountered in process control. Controllers for each process in the batch are then obtained by applying the MIGO design. Having obtained the controller parameters, correlations with normalized process parameters are found by deriving the AMIGO tuning rules. Tables S1 and S2 show these tuning rules for PI and PID controllers in the time and frequency domains. Analysis of these rules can be found in [8]. The main feature of this design method is that it facilitates tradeoffs between robustness and performance. The method thus focuses on load disturbances by maximizing the integral gain and adding a robustness constraint.

TABLE S1 Time-domain AMIGO tuning rules for first-order time delay (FOTD) models. L **represents time delay,** T **is the time constant, and** K_p **is the static gain of the process.** K, T_i, **and** T_d **are proportional gain, integral time, and derivative time parameters of PID controllers.**

Controller	K	T_i	T_d
PI	$\frac{15}{K_p} + \left(0.35 - \frac{LT}{(L+T)^2}\right)\frac{T}{K_pL}$	$0.35L + \frac{13LT^2}{T^2 + 12LT + 7L^2}$	–
PID	$\frac{1}{K_p}\left(0.2 + 0.45\frac{T}{L}\right)$	$\frac{0.4L + 0.8T}{L + 0.1T}L$	$\frac{0.5LT}{0.3L + T}$

TABLE S2 Frequency-domain tuning rules. K_{180} **is the process gain value at frequency** ω_{180}, $T_{180} = (2\pi)/\omega_{180}$ **is the corresponding period, and** $\kappa = K_{180}/K_p$ **is the gain ratio.** K, T_i, **and** T_d **are proportional gain, integral time, and derivative time parameters of PID controllers.**

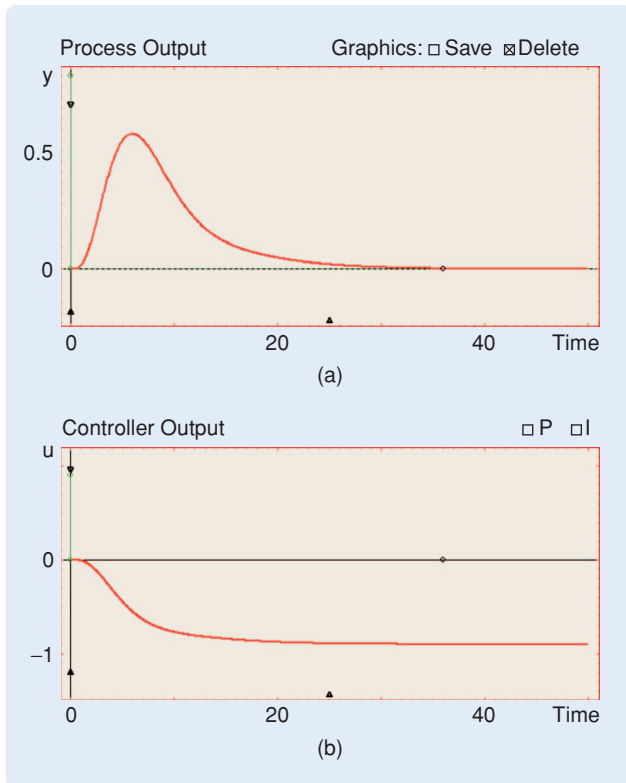| Controller	K	T_i	T_d
PI	$\frac{0.16}{K_{180}}$	$\frac{T_{180}}{1 + 4.5\kappa}$	–
PID	$(0.3 - 0.1\kappa^4)/K_{180}$	$\frac{0.6}{1 + 2\kappa}T_{180}$	$\frac{0.15(1-\kappa)}{1 - 0.95\kappa}T_{180}$

FIGURE 7 Load disturbance response for a PI controller using the approximate M-constrained integral gain optimization (AMIGO) step method. This method enables the compromise described in Figure 6, focusing on load disturbances by maximizing integral gain and adding a robustness constraint. The (a) process output and (b) control signal to load disturbances, respectively, for a PI controller designed using AMIGO and with $K = 0.414$ and $T_i = 2.66$. The slow response compared with Figure 6 corresponds to increased stability margins.
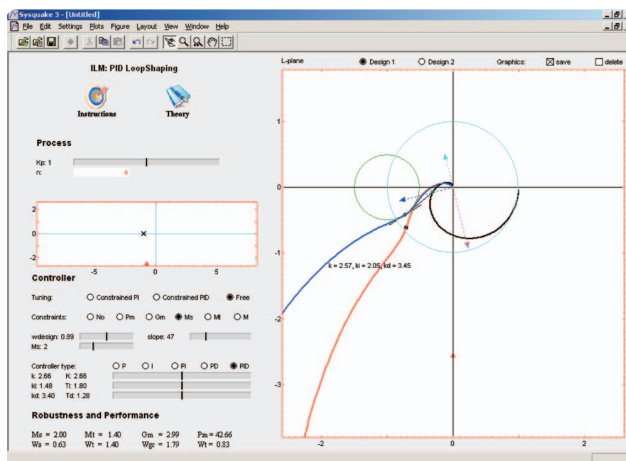
FIGURE 8 The user interface of the module PID Loop Shaping, showing both Free and Constrained PID tuning. The loop transfer function is shown for two designs under the Free design option. Proportional, integral, and derivative action are manipulated directly by drawing the arrows. In the Constrained PID, tuning the target point is constrained to lie on the sensitivity circle.

Several process models are available, and their parameters can be modified using sliders as described in PID Basics. In addition, a free transfer function can be selected (menu option Interactive TF), where poles and zeros can be defined graphically as shown in Figure 8.

Controller

The Controller part of the tool shows the various parameters and properties of PID Loop Shaping to perform loop shaping. The design point of the process transfer function is determined at a specified frequency ω. The design point is shown by a green circle on the L-plane graphic. The corresponding point of the loop transfer function at the frequency ω is called the target point.

The controller used in PID Loop Shaping is parameterized as

$$C(s) = k + \frac{k_i}{s} + k_d s,$$

which yields the loop transfer function

$$L(s) = C(s)P(s) = kP(s) + \left(\frac{k_i}{s} + k_d s\right)P(s).$$

The point on the Nyquist curve of the loop transfer function corresponding to the frequency ω is given by

$$L(i\omega) = kP(i\omega) + i\left(-\frac{k_i}{\omega} + k_d\omega\right)P(i\omega). \quad (4)$$

PID Loop Shaping provides three methods for tuning the parameters to move the process transfer function from the design point to the target point. These methods are listed in the Tuning zone as Free, Constrained PI, and Constrained PID. Free tuning allows an unconstrained loop to be shaped by dragging on the control parameters. Constrained PI and Constrained PID permit the calculation of the controller parameters based on some constraints on the target point. That is, the focus can be placed on how the loop transfer function changes when controller parameters are modified, which reveals the parameter values required to obtain a given shape of the loop transfer function. For PI and PD control the mapping is uniquely given by one point. For PID control it is also possible to obtain an arbitrary slope ϑ of the loop transfer function at the target point. When the Free tuning option is selected, sliders are used to modify the controller gains k, k_i, and k_d, as shown in Figure 8. The controller gains can also be changed by dragging arrows, as illustrated in the same figure. From (4), the proportional gain changes $L(i\omega)$ in the direction of $P(i\omega)$, the integral gain k_i changes $L(i\omega)$ in the direction of $-iP(i\omega)$, and the derivative gain k_d changes $L(i\omega)$ in the direction of $iP(i\omega)$.

For the Constrained PI and Constrained PID tuning options, the target point can be limited to move on the unit circle, the sensitivity circles, or the real axis. In this

way loop shaping is enabled with specifications on gain and phase margins or on the sensitivities. In the case of Constrained PI it is necessary to find controller gains providing the desired target point. Dividing (4) by $P(i\omega)$ and separating the real and imaginary parts gives

$$k = \Re\left(\frac{L(i\omega)}{P(i\omega)}\right), \tag{5}$$

$$-\frac{k_i}{\omega} + k_d\omega = \Im\left(\frac{L(i\omega)}{P(i\omega)}\right) = A(\omega). \tag{6}$$

With $k_d = 0$, (5) and (6) yield the two parameters of the PI controller.

An additional condition is required for the Constrained PID tuning option. Hence, it is observed that

$$\begin{aligned}
L'(s) &= C'(s)P(s) + C(s)P'(s) \\
&= C'(s)P(s) + \frac{L(s)P'(s)}{P(s)} \\
&= \left(-\frac{k_i}{s^2} + k_d\right)P(s) + \frac{L(s)P'(s)}{P(s)}.
\end{aligned} \tag{7}$$

The slope of the Nyquist curve is then given by

$$iL'(i\omega) = i\left(\frac{k_i}{\omega^2} + k_d\right)P(i\omega) + iC(i\omega)P'(i\omega). \tag{8}$$

The complex number represented by (8) has the phase angle ϑ if

$$\Im(iL'(i\omega)e^{-i\vartheta}) = 0. \tag{9}$$

Results (7)–(9) imply that

$$\frac{k_i}{\omega^2} + k_d = \frac{\Re\left(L(i\omega)\frac{P'(i\omega)}{P(i\omega)}e^{-i\vartheta}\right)}{\Re(P(i\omega)e^{-i\vartheta})} = B(\omega). \tag{10}$$

Combining (10) with (5)–(6) gives the controller parameters

$$k_i = -\omega A(\omega) + \omega^2 B(\omega), \tag{11}$$

$$k_d = \frac{A(\omega)}{\omega} + B(\omega), \tag{12}$$

where $A(\omega)$ and $B(\omega)$ are given by (6) and (10), respectively.

The design frequency ω can be chosen using the slider wdesign or graphically by dragging the green circle on the process Nyquist curve (black curve in Figure 8). The target point on the Nyquist plot and its slope can be dragged graphically. The slope can also be changed using the slider slope. Furthermore, it is possible to constrain the target point using the Constraints radio buttons to the unit circle

(Pm), the negative real axis (Gm), circles representing constant sensitivity (Ms), constant complementary sensitivity (Mt), or constant sensitivity combinations (M). When sensitivity constraints are active, the associated circles are drawn in the L-plane plot, and sliders can be used to modify their values. The circles are defined in Table 1.

Figure 8 illustrates designs for two PID controllers and a given sensitivity. The target point is moved to the sensitivity circle, and the slope is adjusted so that the Nyquist curve is outside the sensitivity circle. The red design shows a PID controller using Free tuning, while the blue design shows a Constrained PID tuning. Specifications that cannot be reached are indicated in the tool by giving the integral or derivative gain negative values in these cases.

Robustness and Performance Parameters

Robustness and Performance parameters are displayed on the screen below the controller parameters (Figure 8), and these parameters characterize robustness and performance in the same manner as in PID Basics. The values are maximal sensitivity (Ms), sensitivity-crossover frequency (Ws), maximal complementary sensitivity (Mt), complementary sensitivity-crossover frequency (Wt), gain margin (Gm), gain-crossover frequency (Wgc), phase margin (Pm), and phase-crossover frequency (Wpc).

L-Plane Graphic

The L-plane graphic is given in the right-hand side of the PID Loop Shaping menu, as shown in Figure 8. This graphic contains the Nyquist plots of the process transfer function $P(s)$ in black and the loop transfer functions $L(s) = P(s)C(s)$ in red. Three different views can be shown depending on the tuning options. Figure 9 shows two views, the left one for Free tuning and the right one for Constrained PID tuning. A third view is shown in Figure 8, where two designs are shown simultaneously. The design and target points can be modified interactively on this graphic. The design point is shown in green on the Nyquist curve of the process. The target point is represented in light green in the case of Free tuning and in black for constrained tuning, as shown in Figure 8. The slope of the target point can also be changed

TABLE 1 Sensitivity circles. This table describes the center and radius of circles that define the loci for constant sensitivity M_s, constant complementary sensitivity M_t, constant mixed sensitivity, and equal sensitivities $M = M_s = M_t$ [8].

Contour	Center	Radius
M_s-circle	-1	$1/M_s$
M_t-circle	$-\frac{M_t^2}{M_t^2-1}$	$\frac{M_t}{M_t^2-1}$
M-circle	$-\frac{x_1+x_2}{2}$	$\frac{x_1-x_2}{2}$
$x_1 = \max\left(\frac{M_s+1}{M_s}, \frac{M_t}{M_t-1}\right)$		$x_2 = \max\left(\frac{M_s-1}{M_s}, \frac{M_t}{M_t+1}\right)$

interactively. For Free tuning, the controller gains are shown as arrows in the Nyquist plot. The controller gains can be modified interactively by dragging the ends of the arrows. Figures 8 and 9 show examples of these arrows. The scale of the graphic can be changed using the red triangle located at the bottom of the vertical axis.

As noted above, it is possible to impose constraints on the target point. The graphical representation of the target point is modified depending on the constraint selected, restricting

(a)

(b)
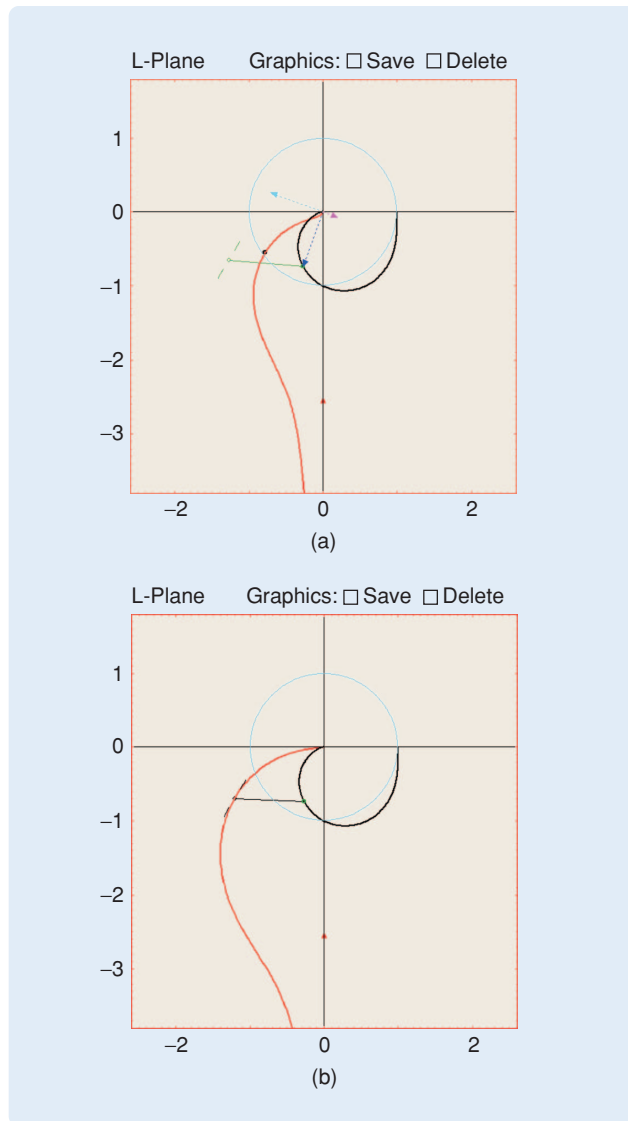
FIGURE 9 The L-plane graphic. The Nyquist plots of the process transfer function $P(s)$ (black line) and the loop transfer function $L(s) = P(s)C(s)$ (red line) are shown. (a) An example of the Free tuning design. The controller gains can be changed by dragging arrows, the proportional gain changes $L(i\omega)$ in the direction of $P(i\omega)$ (blue arrow), the integral gain k_i changes $L(i\omega)$ in the direction of $-iP(i\omega)$ (cyan arrow), and the derivative gain k_d changes $L(i\omega)$ in the direction of $iP(i\omega)$ (magenta arrow). (b) An example of the Constrained PID tuning design. In this case, once the user moves the target point (black circle), the controller parameters are calculated using (5)–(12).

its value based on its meaning. Options save and delete can be found above the L-plane graphic. These options have the same meaning as in PID Basics, making it possible to save designs to perform comparisons. Once the save option is active, two pictures appear, one of which shows the current design in red while the other shows the current design in blue (see Figure 8). Modifications of the controller parameters affect the current (active) design, which can be changed using the options Design 1 and Design 2, which appear on the top of the L-plane graphic. Once a design is chosen, the associated curve is switched to red, and the controller zone is modified based on that design. The controller gain values can be seen by moving the cursor on the curves.

Settings Menu

The Settings menu, which is available in the main menu of PID Loop Shaping, is divided into four groups, following the same structure as in PID Basics. The first entry, called Process Transfer Function, is used to choose between several predefined transfer functions or to include an user-specified transfer function through two options. The String TF option allows a transfer function to be entered symbolically. For instance, $P(s) = 1/\cosh\sqrt{s}$ can be represented as P='1/cosh(sqrt(s))'. Results can be stored and recalled using the Load/Save menu. The option Save Report can be used to save all essential data in text format, which is useful for documenting results. Specific values for control parameters can be entered with Parameters menu option. As in PID Basics, the last menu option (Examples Advanced PID Book) allows loading examples from [8].

Examples

Some of the capabilities of PID Loop Shaping are illustrated by the following examples.

Effect of Controller Parameters

The purpose of this example is to illustrate how the Nyquist plot of the loop transfer function changes when the controller parameters are modified.

Consider the process $P(s) = 1/(s+1)^4$. When a P controller is used, the proportional gain changes the loop transfer function $L(i\omega) = kP(i\omega)$ in the direction of $P(i\omega)$. Figure 10(a) shows the effect of modifying $L(i\omega)$ using a P controller with gain $k = 2$ (blue curve) and $k = 2.6$ (red curve). These curves show how the proportional gain modifies the Nyquist plot of the process (black curve) at the frequency ω (green circle on the black curve) in the direction of $P(i\omega)$. Figure 10(b) shows the same study for an I-controller with $k_i = 1$ (red curve) and $k_i = 0.6$ (blue curve). It can be seen that the integral gain k_i changes $L(i\omega)$ in the direction $-iP(i\omega)$. The derivative gain has the same effect in the direction of $iP(i\omega)$. When a PI or PD controller is used, the compensated point at the frequency ω is calculated as the sum of two vectors, namely, the proportional vector and the integral or derivative vector. Examples of this capability are

shown in Figure 10(c) and (d), where the process is controlled by a PI controller ($k = 2.3$ and $k_i = 0.7$) and a PD controller ($k = 2.1$ and $k_d = 3.35$), respectively.

Simple exercises can be used to provide training in loop shaping. For instance, with the above process, it is instructive to calculate the gain for a proportional controller for which the closed-loop system changes from stable to unstable. Before using PID Loop Shaping, the result can be calculated analytically, which yields

$$\angle L(i\omega) = \angle C(i\omega)P(i\omega) = -180,$$
$$\angle k \frac{1}{(i\omega + 1)^4} = -180, \quad \omega = 1.$$
$$|L(i\omega)| = |C(i\omega)P(i\omega)| = |-1 + 0i|,$$
$$\left| k \frac{1}{(i\omega + 1)^4} \right| = -1, \quad k = 4.$$

PID Loop Shaping can be used to verify the result interactively, as shown in Figure 11. This exercise challenges students and encourages them to make observations while relating theory to images to develop a broader and deeper understanding.

On the other hand, free interactive designs can also be performed to compare the results with other design methods. For instance, PID Loop Shaping can be used to design a PID controller interactively for the process $P(s) = 1/(s + 1)^4$, where the maximal sensitivity value M_s must be less than 1.5. A PID controller that satisfies this constraint is obtained when $k = 0.92$, $T_i = 1.8$, $k_i = 0.5$, $T_d = 1.03$, and $k_d = 0.95$. The AMIGO-frequency method can also be used for design, and the results can be compared. The resulting controller is given by $k = 1.2$, $T_i = 2.48$, $k_i = 0.48$, $T_d = 0.93$, and $k_d = 1.12$. Figure 12 shows the Nyquist plots and time responses using PID Basics for both designs, in blue for the free PID controller and in red for the AMIGO method. The resulting values of M_s are 1.49 for free PID and 1.46 for the AMIGO method.
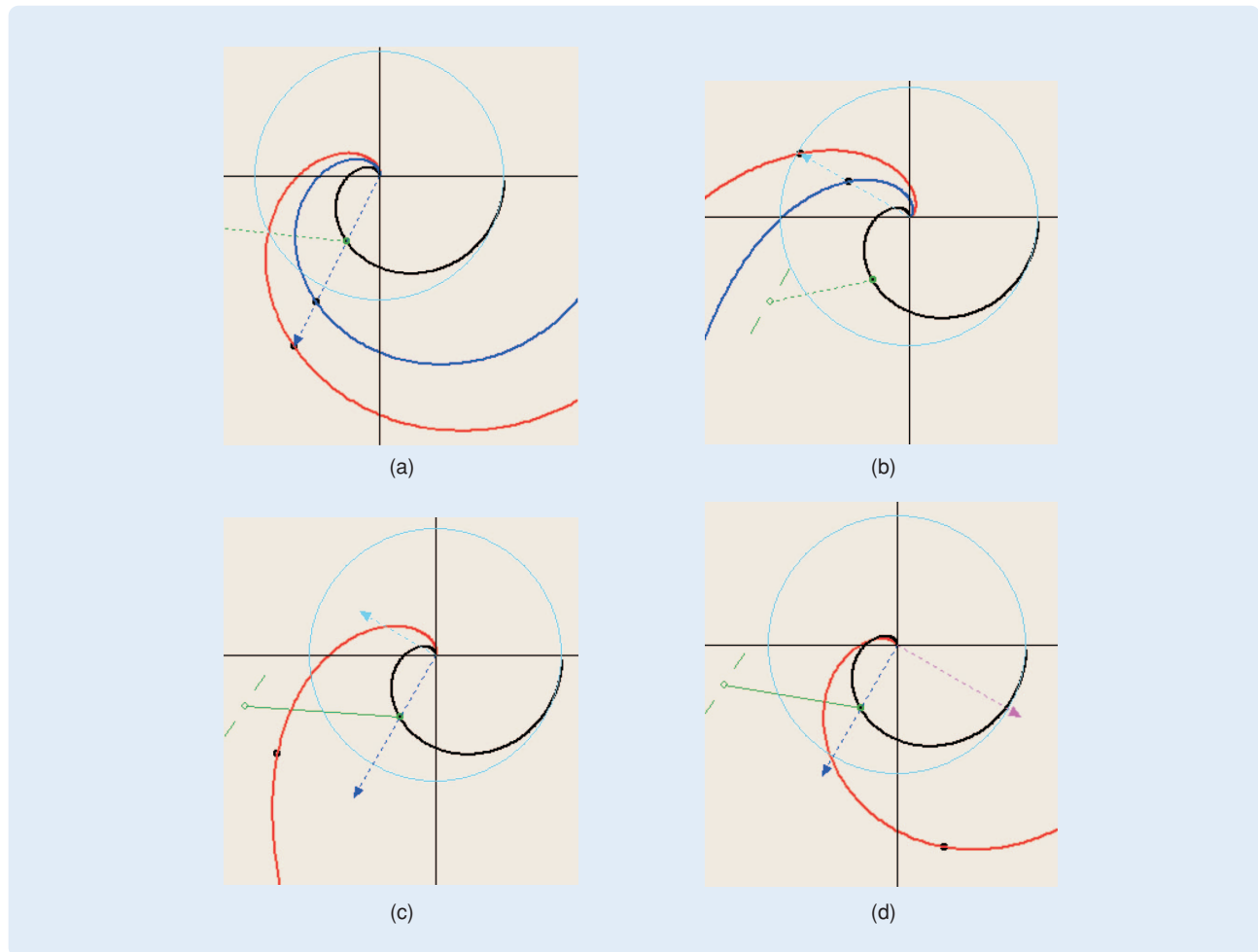
FIGURE 10 Nyquist plot modifications depending on the controller type. (a) P controller, (b) I controller, (c) PI controller, and (d) PD controller. (a) The modification of $L(i\omega)$ in the direction of $P(i\omega)$ using a P controller with gain $k = 2$ (blue curve) and $k = 2.6$ (red curve). The same study for an I-controller is shown in (b) with $k_i = 1$ (red curve) and $k_i = 0.6$ (blue curve), where $L(i\omega)$ is modified in the direction $-i P(i\omega)$. (c), (d) PI or PD controllers are used, respectively. In these cases, the compensated point at the frequency ω is calculated as the sum of two vectors, namely the proportional vector and the integral or derivative vector.

Effect of the Target Point

The target point on the Nyquist plot can be reached using an unconstrained design by selecting the Free option. The controller gains are interactively adjusted as shown in the free tuning example. Another approach is to use (5)–(12), where the controller gains are calculated after the target point is defined. As discussed above, the target point can be fixed or constrained in various ways, either at any point, to specific values for phase margin and gain margin, or to maximal values of the sensitivity functions. Figure 13 shows an example in which the target point is set to the point $-0.5 - 0.5i$. Two constrained designs are shown for the design frequency $\omega = 0.6$ rad/s. The red curve represents a system compensated by a constrained PID with $k = 1.32$, $k_i = 1.02$, and $k_d = 2.15$, while the blue curve represents a constrained PI with $k = 1.32$ and $k_i = 0.15$. Although both controllers reach the target point, better results are obtained for the PID controller because the slope can be freely adjusted (the value for this example is $\vartheta = 22$). The PID controller provides better robustness properties with $M_s = 1.45$, $k_i = 1.02$, $G_m = 5.32$, and $P_m = 40.15$, versus a PI controller with $M_s = 1.83$, $k_i = 0.15$, $G_m = 2.69$, and $P_m = 75.77$.

Similar examples can be used to restrict the target point for phase margin, gain margin, or maximal values of the sensitivity functions. Figure 14(a) shows an example where a combined sensitivity constraint is required for $M_s \leq 2$ and $M_t \leq 2$. This constraint is fulfilled in two different ways, namely, by using Constrained PID (red curve) and Constrained PI (blue curve). Another example combining sensitivity function and gain margin constraints is shown in Figure 14(b), with the specification that the gain margin be equal to three and $M_s \leq 2$. These specifications are established by maximizing the integral gain k_i. Hence, the constraint gain margin is chosen, and the target point is located in such a way that $G_m = 3$. Then, a Constrained PID controller is selected, where the design point and the slope are modified until $M_s \leq 2$ and the integral gain is maximized. The final controller is given by $k = 1.38$, $k_i = 0.52$, and $k_d = 0.54$ for $\omega = 1.02$ and $\vartheta = 32$.

(a)

(b)
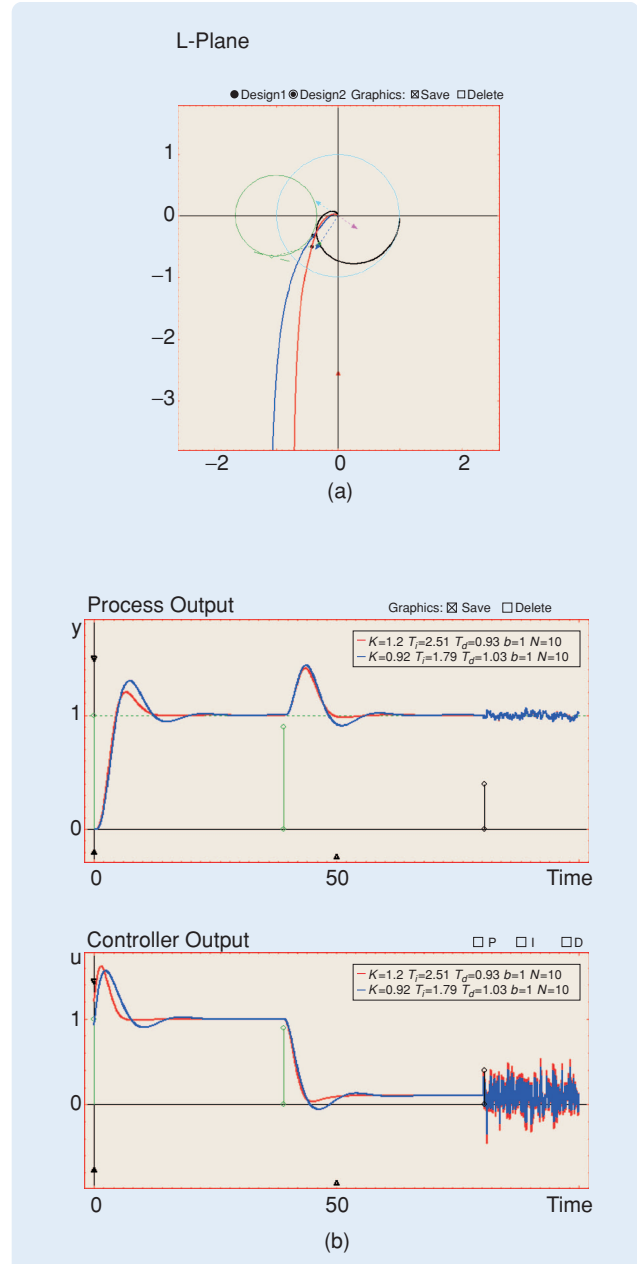
FIGURE 12 Example of loop shaping with $M_s < 1.5$. PID Loop Shaping can be used to compare various designs. In this figure, (a) Nyquist plots and (b) time-domain responses generated with PID Basics are shown to compare an unconstrained design ($k = 0.92$, $T_i = 1.8$, $k_i = 0.5$, $T_d = 1.03$, and $k_d = 0.95$) with an alternative design developed using the AMIGO-frequency method ($k = 1.2$, $T_i = 2.48$, $k_i = 0.48$, $T_d = 0.93$, and $k_d = 1.12$).

FIGURE 11 Stability limit on the critical point $-1 + 0i$. A typical example for presenting loop shaping is to search for the lowest gain that makes the system unstable. This task can be interactively performed with PID Loop Shaping as shown in this figure.

The Derivative Cliff

We again consider the process transfer function $P(s) = 1/(s+1)^4$. It is desirable to maximize the integral gain k_i subject to the robustness constraint $M_s \leq 1.4$. The resulting controller has the parameters $k = 0.925$, $k_i = 0.9$, and $k_d = 2.86$, where the Nyquist plot of the loop transfer function is shown in red in Figure 15(a). It can be seen that the Nyquist curve has a loop, called a derivative cliff. As explained in [8], this feature, which is due to excessive controller phase lead, results from having a PID controller with complex poles, which occurs when $T_i < 4T_d$. In this example the relation is $T_i = 0.33T_d$. Figure 15(b) shows, in red, the time response of the controller, which yields oscillatory outputs. For comparison, the results for a controller with $T_i = 4T_d$ are shown in blue in Figure 15(a) and (b) with the controller parameters $k = 1.1$, $k_i = 0.36$, and $k_d = 0.9$. The responses for this controller are improved, despite larger overshoot in response to load disturbances. This example is available in the Settings menu of PID Loop Shaping.

PID WINDUP

The purpose of the PID Windup module is to facilitate understanding of integral windup and a method for compensating it [8]. For a control system with a wide range of operating conditions, it may happen that the control variable reaches the actuator limits. When this situation occurs in loops using a controller with integral action, the feedback loop is broken and the integral may reach large val-

ues, maintaining the control signal saturated for a long time, resulting in large overshoot, and undesirable transients. This problem is known as windup phenomenon [8].

Windup can be avoided in different ways. Back calculation and tracking [8] is illustrated in the block diagram in Figure 16. The system remains unchanged when the saturation is not active. However, when saturation occurs, the integral term in the controller is modified until the control signal is out of the saturation limit. This modification is not performed instantaneously but dynamically with a time constant T_t called the tracking time constant [8].

The module PID Windup shows process outputs and control signals for unlimited control signals, limited
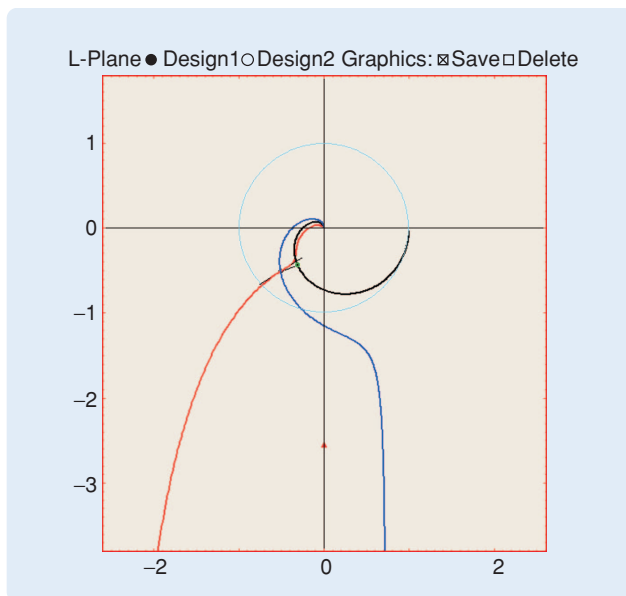
FIGURE 13 Example of a constrained design with a target point of $-0.5 - 0.5i$. The target point can be constrained to reach arbitrary specifications. Once a point is constrained, the controller parameters are automatically calculated. This plot shows PI ($k = 1.32$, $k_i = 0.15$) and PID ($k = 1.32$, $k_i = 1.02$, $k_d = 2.15$) controllers, both reaching the target point. The PID controller provides better results due to the use of the slope as an allowable third degree of freedom as described in (11) and (12), where the slope ϑ takes the value 22.
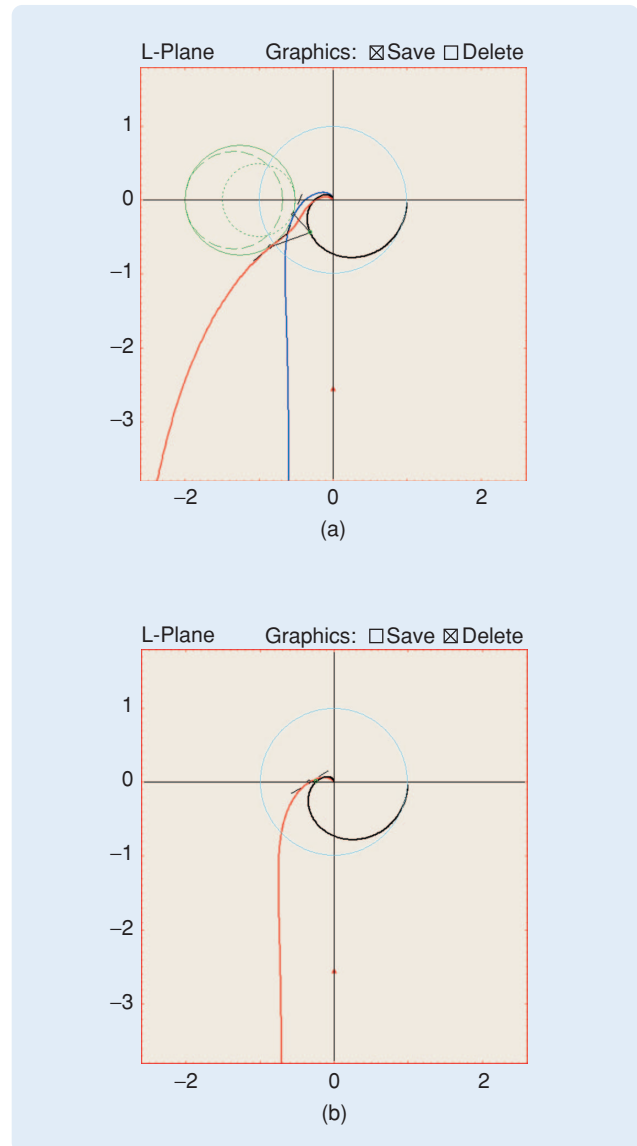
FIGURE 14 Example of a constrained design with sensitivity and gain-margin constraints. These plots show an example where the target point is constrained to reach specified values for the (a) combined sensitivity functions with $M_s \leq 2$ and $M_t \leq 2$, and (b) gain margin with limited sensitivity values with $Gm = 3$ and $M_s \leq 2$.

control signals without antiwindup, and limited control signals with antiwindup. The user interface is shown in Figure 17. Process models and controller parameters can be selected in the same way as in the other modules. The saturation limits of the control signal can be determined either by entering the values or by dragging the lines in the saturation scheme.
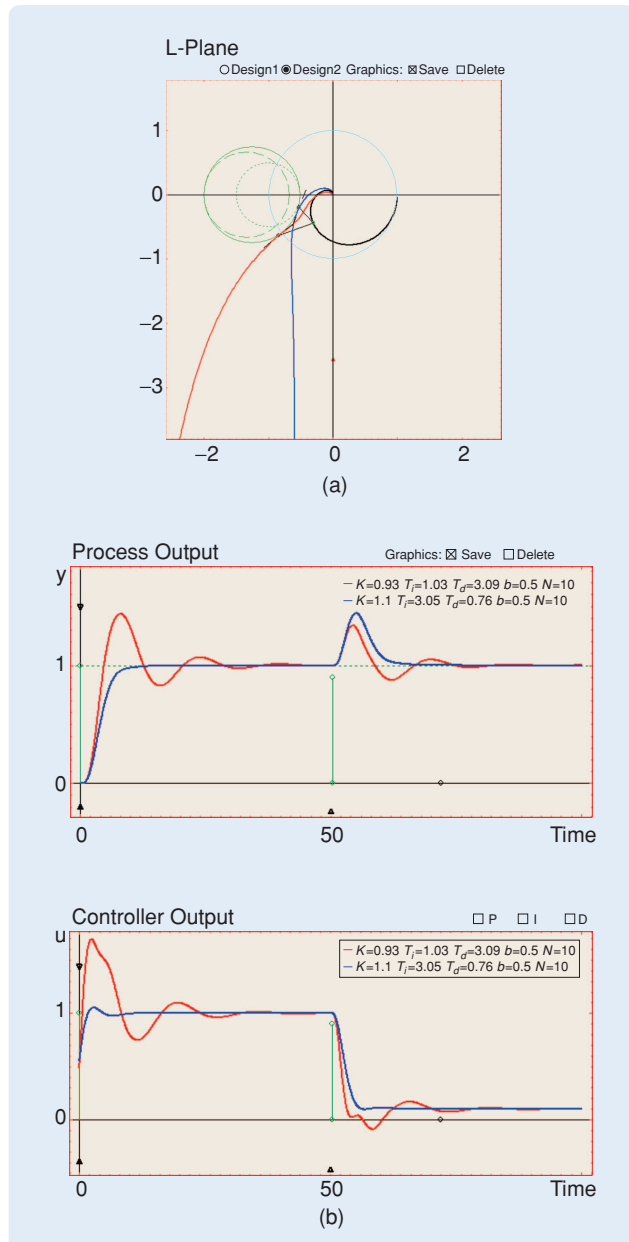

FIGURE 15 Derivative cliff example. (a) Nyquist plot and (b) time-domain responses. This example shows that optimization of k_i, which is aimed at fulfilling robustness specifications, can provide controllers with excessive phase lead, as represented by the loop in the red curve. This behavior is a consequence of the presence of complex zeros due to $T_i < 4T_d$. The same example is shown in blue for $T_i = 4T_d$, where this problem is avoided [8].

The Interactive Tool

We now describe the main aspects of PID Windup.

Process

The Process area is similar to that described in PID Basics and PID Loop Shaping. The time delay is modified using a slider instead of a text edit, so that the time delay effect on the antiwindup mechanism can be analyzed.

Controller

The Controller area contains information about the controller parameters and actuator saturation. Three kinds of
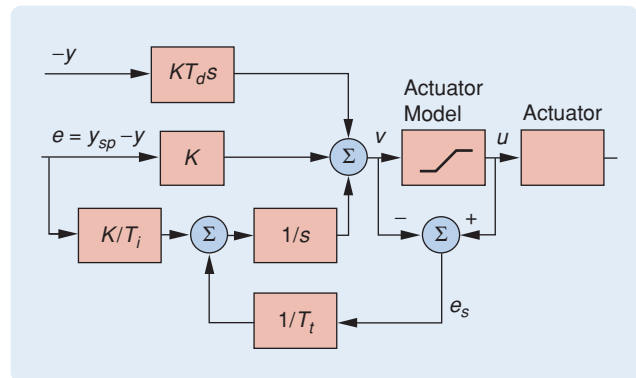
FIGURE 16 PID controller with antiwindup scheme, where K is the controller proportional gain, T_i is the controller integral time, T_d is the controller derivative time, y_{sp} is the setpoint, y is the process output, e is the tracking error, v is the controller output, u is the saturated controller output, and e_s is the difference between the controller output v and the saturated controller output u. In this scheme the control signal remains unconstrained when the saturation is not active. When saturation occurs, the integral control action is modified until the control signal is out of the saturation limit. The modification of the integral element is performed dynamically by adjusting the tracking time constant T_t [8].
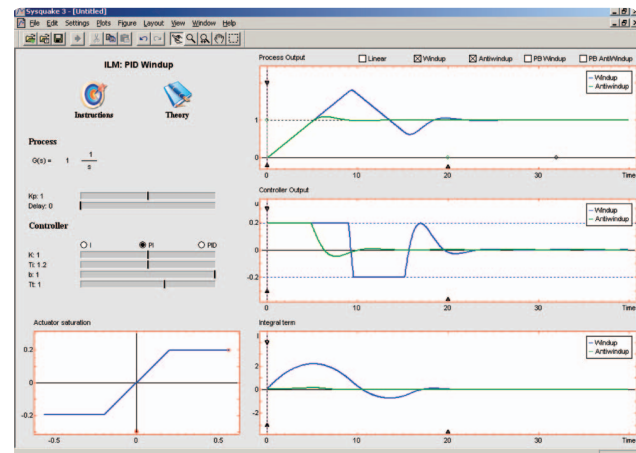
FIGURE 17 The user interface of the module PID Windup, showing the windup phenomenon and application of the antiwindup technique. Several graphical elements are used to interactively analyze typical problems and solutions associated with windup. The example shown in the figure illustrates the windup phenomenon (in blue) and the result of applying the antiwindup technique (in green).

controllers with integral action can be selected (I, PI, PID), where several sliders are available to change the controller parameters, including the tracking time constant Tt. A saturation graphic is also available in this zone. The Actuator Saturation graphic allows the saturation limits to be determined by dragging the small red circle located on the upper saturation value. In this graphic, a symmetric saturation is selected for pedagogical purposes.

Graphics

Time responses for process output, control signal, and integral action are available in three graphics, namely, Process Output, Controller Output, and Integral Term. In the same way as in PID Basics, multiple interactive graphical elements can be used to change the setpoint, load disturbance, measurement noise, or horizontal and vertical scales (see Figure 17). These graphics can simultaneously represent the controlled system in linear, nonlinear with windup, and nonlinear with antiwindup modes. These representations can be configured using the checkboxes located above the Process Output graphic. For instance, Figure 17 shows an example containing the nonlinear with windup and nonlinear with antiwindup modes.

The dotted pink vertical line in Figure 18 is helpful for comparing the outputs of the different plots at the same time instant. The saturation limits can be altered using the

dotted blue horizontal lines available in the Controller Output graphic (see Figure 17).

The notion of proportional band is useful for understanding the windup effect, and is included in PID Windup. The proportional band is defined as the range of process outputs such that the controller output is in the linear range $[y_{min}, y_{max}]$. For a PI controller, the proportional band is limited by

$$y_{min} = by_{sp} + \frac{I - u_{max}}{K}, \tag{13}$$

$$y_{max} = by_{sp} + \frac{I - u_{min}}{K}, \tag{14}$$

where I is the integral term of the controller, and u_{max} and u_{min} are the control signal limits.

Expressions (13) and (14) hold for PID control when the proportional band is defined as the band where the predicted output $y_p = y + T_d(dy/dt)$ is in the proportional band $[y_{min}, y_{max}]$. The proportional band has the width $(u_{max} - u_{min})/K$, and is centered around $by_{sp} + I/K - (u_{max} + u_{min})/(2K)$.

Two additional checkboxes called PB Windup and PB Antiwindup, appear near the top of plot Process Output. The activation of these options shows the proportional bands for the windup and antiwindup cases in the Process Output graphic. The proportional bands are shown as dotted green and blue curves, respectively, as shown in Figure 18.
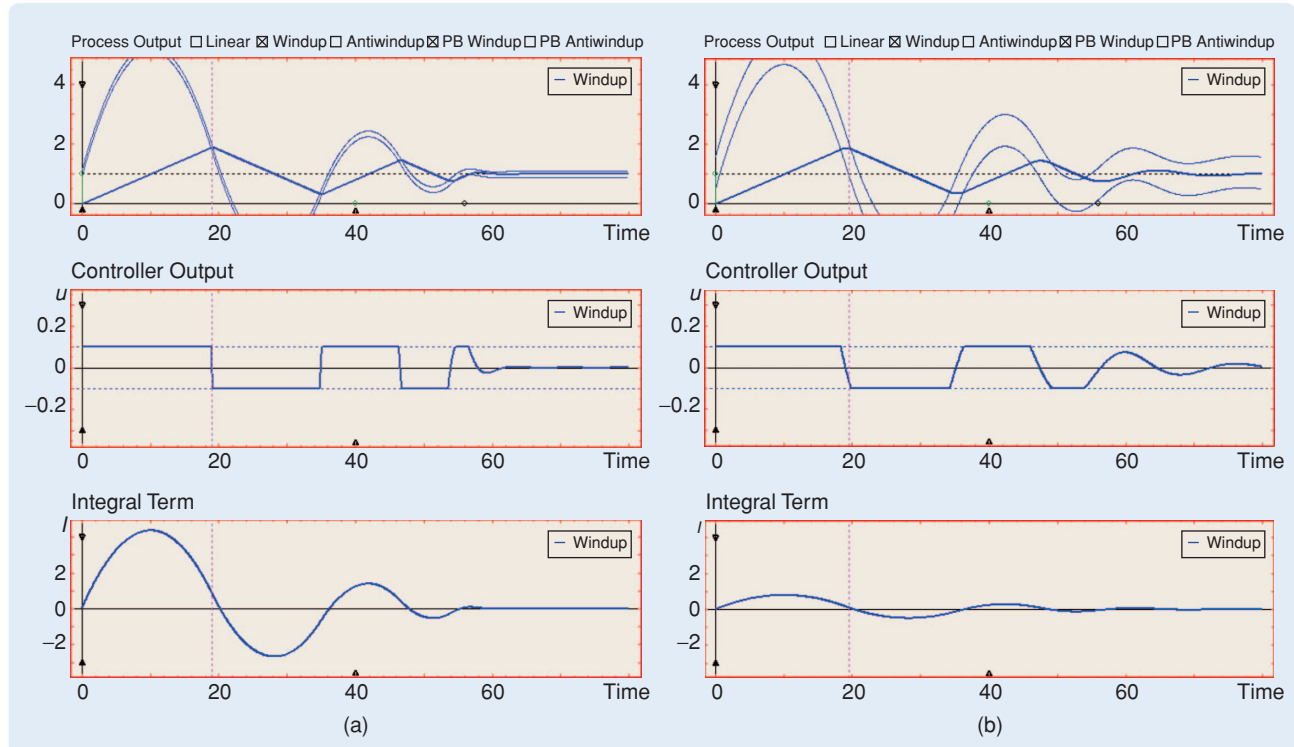
FIGURE 18 Example of the windup phenomenon with proportional band for (a) $K = 1$ and (b) $K = 0.4$. In [8] the notion of proportional band is described as being a useful tool for understanding the effects of windup. The proportional band is an interval such that the actuator does not saturate when the instantaneous value of the process output or its predicted value is inside this band. These plots show two examples demonstrating how the control signal is saturated when the process output is inside the band shown in blue. The interactive pink line of the graphics can be used to test this idea.

Settings Menu

The Settings menu has the same structure as in PID Basics and PID Loop Shaping. The process transfer function can be chosen from the entry Process Transfer Functions, and numerical values of the parameters can be introduced using Controller Parameters. Essential data and results can be saved and recalled using the Load/Save menu options. The menu selection Simulation makes it possible to choose the simulation time and activate the Sweep option, which can be used to show the results for several values of the tracking time constant. Several examples from [8] can be loaded from the Examples entry.

Examples

The following examples illustrate properties of the PID Windup module.

Understanding the Windup Phenomenon

Windup can be studied using the first entry from the Examples option menu. This example from [8] uses the pure integrator process $P(s) = 1/s$ controlled by a PI controller with parameters $K = 1$, $T_i = 1.2$, and $b = 1$, and with the control signal limited to ± 0.2. Figure 17 shows the time responses for this example. The control signal is saturated from $t = 0$. The process output and the integral term increase while the control error is positive. Once the

process output exceeds the setpoint, the control error becomes negative, however the control signal remains saturated due to the large value of the integral term. The time responses are shown in Figure 17.

The proportional band can be drawn in this example using the PB Windup checkbox shown in Figure 18(a). Using the vertical line, the user can see that the process output remains inside the band while the control signal is working in linear mode and outside the proportional band when the control signal is saturated. Large controller gains provide narrow proportional bands, with more energetic control signals and therefore longer saturation times, while small controller gains give wider proportional bands. Figure 18(b) illustrates this effect, where the proportional controller gain is reduced to 0.4, producing a wider proportional band.

Antiwindup

The process $P(s) = 1/s$ is also useful for visualizing the antiwindup technique. The same controller parameters, namely, $K = 1$, $T_i = 1.2$, $b = 1$, are used, and the tracking time constant is set to $T_t = 1$. Figure 19(a) shows the responses for both cases control with and without antiwindup. The system with antiwindup remains in saturation for only a short period of time, with the magnitude of the integral term considerably reduced. The proportional
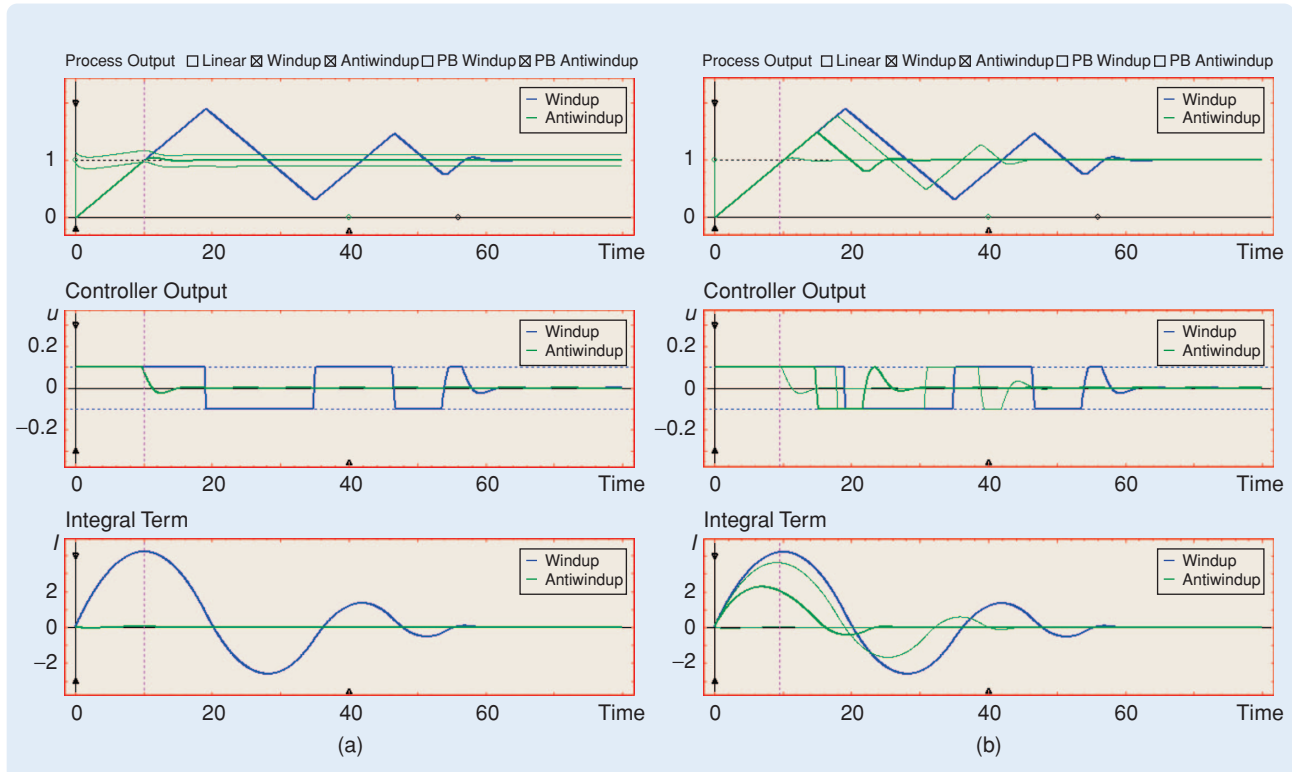
FIGURE 19 Example of the effect of the tracking time constant T_t on in the antiwindup technique. (a) Antiwindup and (b) effect of T_t. These plots show the results of applying the antiwindup technique to the example shown in Figure 17. The integral signal is considerably reduced, allowing the control signal to remain in saturation during a shorter period of time. The proportional band for the antiwindup technique is shown in green. The process output remains inside the band most of the time.

band for the PI controller with antiwindup is shown in the same figure. It can be seen that the proportional band is wider than for PI without antiwindup [Figure 19(a)], where the process output remains most of the time. The effect of the tracking time constant is illustrated in Figure 19(b) for $T_t = 0.1, 10, 50$. In this scenario, the Sweep menu option is used. High values of T_t make the antiwindup too slow to be effective, while low values reset the integral term quickly with improved results. It may thus seem advantageous to always have small values of T_t. However, the next example shows some situations where this choice is not advisable.

The Tracking Time Constant

The tracking time constant is an essential parameter because it determines the reset rate for the integral term of the controller. It may seem advantageous to have a small value for this constant. However, measurement errors may accidentally reset the integral term when the tracking time constant is too small. The following example illustrates this phenomenon, when a measurement error occurs in the form of a short pulse. The transfer function of the process is

$$P(s) = \frac{1}{(0.5s + 1)^2},$$

and the controller is a PID controller with $K = 3.5$, $T_i = 0.52$, $T_d = 0.14$, $N_d = 10$, $b = 1$, and $T_t = 1$.

Figure 20(a) shows the control results. A large transient appears after the pulse, and the integral term is excessively reduced.

Various rules are suggested in [8] for choosing the tracking time constant. One choice is $T_t = (T_i + T_d)/2$. Figure 20(b) shows an example with $T_t = (T_i + T_d)/2 = 0.33$, where the response is considerably improved.

CONCLUSIONS

In this work a set of interactive modules that comprise ILM-PID is presented to support the teaching and learning of basic automatic control concepts. These tools are intended mainly to include interactivity in the visual content of [8]. The modules focus on PID control, studying feedback fundamentals from the standpoint of the time and frequency domains, including robustness issues, measurement of noise filtering, load-disturbance rejection, and windup phenomenon.

The importance of interactivity in automatic control education has been shown in the context of teaching and learning. In the authors' experience, interactivity offers excellent support to education and learning by enhancing the motivation and participation of future
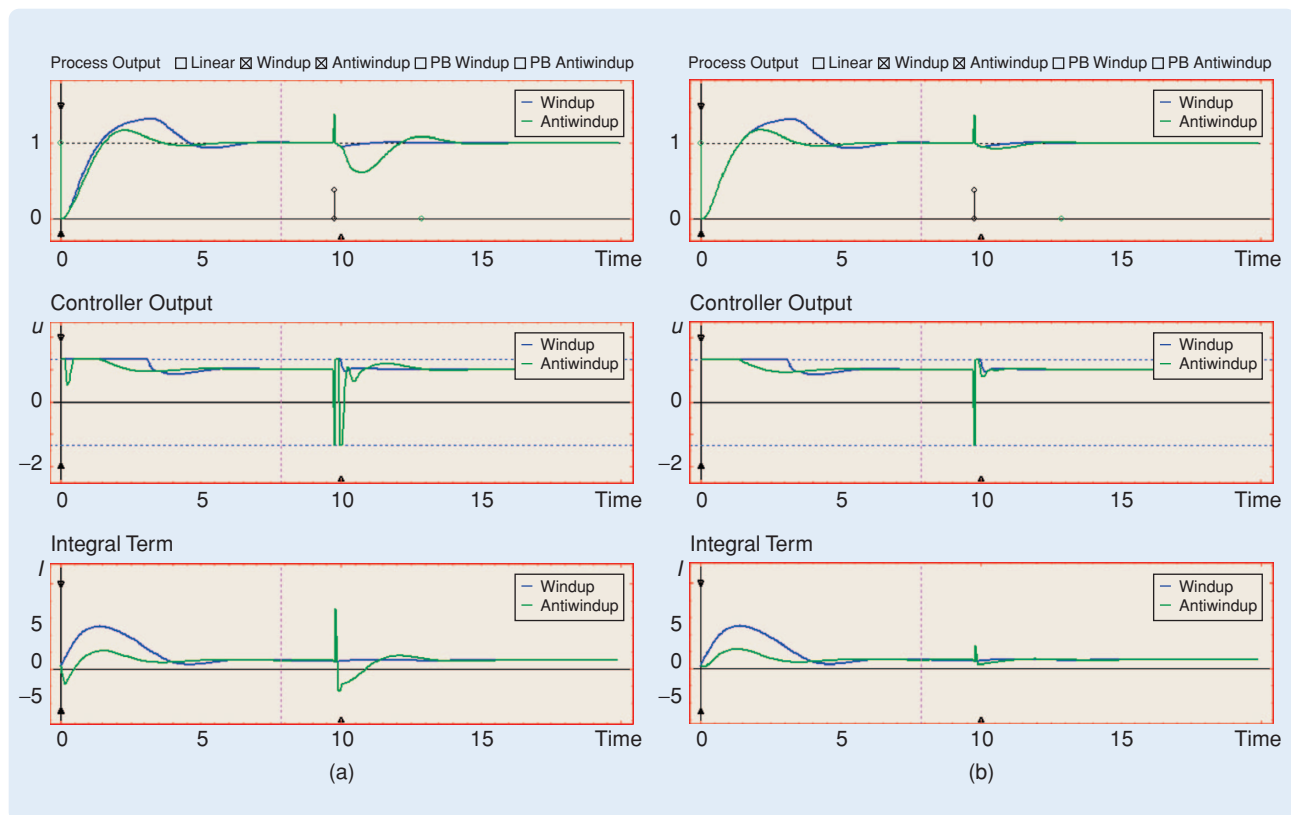
FIGURE 20 Tuning the tracking time. (a) Reset by measurement noise and (b) tuning using rules. (a) illustrates the disadvantage of using a short tracking time constant. The short pulse disturbance at time $t = 10$ results in excessive reduction of the integral term and a large disturbance in the process output. In (b) the choice is $T_t = (T_i + T_d)/2$.

engineers. The interactive learning modules developed in this work are freely available from the authors [7] to test these interactive features in control education and professional training.

REFERENCES

[1] M. Johansson, M. Gäfvert, and K.J. Åström, "Interactive tools for education in automatic control," *IEEE Contr. Syst. Mag.*, vol. 18, no. 3, pp. 33–40, 1998.

[2] J.L. Guzmán, M. Berenguel, and S. Dormido, "Interactive teaching of constrained generalized predictive control," *IEEE Contr. Syst. Mag.*, vol. 25, no. 2, pp. 79–85, 2005.

[3] J. Sánchez, S. Dormido, and F. Esquembre, "The learning of control concepts using interactive tools," *Comput. Appl. Eng. Educ.*, vol. 13, no. 1, pp. 84–98, 2005.

[4] J.L. Guzmán, "Interactive control system design," Ph.D. dissertation, University of Almería, Spain [Online]. Available: http://www.ual.es/~joguzman/PhdThesisGuzman.pdf

[5] Y. Piguet, *SysQuake 3 User Manual*. Lausanne, Switzerland: Calerga Sárl, 2004 [Online]. Available: http://www.calerga.com

[6] J.L. Guzmán, K.J. Åström, S. Dormido, T. Hägglund, and Y. Piguet, "Interactive learning modules for PID control," in *Proc. 7th IFAC Symp. Advances in Control Education*, 2006, Madrid, Spain [Online]. Available: http://aer.ual.es/ilm/

[7] ILM Web site, *Interactive Learning Modules*, 2005 [Online]. Available: http://www.calerga.com/contrib/1/index.html

[8] K.J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC : Instrum. Soc. Amer. 2005.

[9] F.G. Shinskey, *Process-Control Systems. Application, Design, and Tuning*. New York: McGraw Hill, 1996.

[10] T. Hägglund and K.J. Åström, "Revisiting the Ziegler-Nichols tuning rules for PI control," *Asian J. Contr.*, vol. 4, no. 4, pp. 364–380, 2002.

[11] T. Hägglund and K.J. Åström, "Revisiting the Ziegler-Nichols step response method for PID control," *J. Process Contr.*, vol. 14, no. 6, pp. 635–650, 2004.

[12] T. Hägglund and K.J. Åström, "Revisiting the Ziegler-Nichols tuning rules for PI control—Part II, the frequency response method," *Asian J. Contr.*, vol. 6, no. 4, pp. 469–482, 2004.

[13] K.J. Åström, H. Panagopoulos, and T. Hägglund, "Design of PI controllers based on non-convex optimization," *Automatica*, vol. 34, no. 4, pp. 585–601, 1998.

AUTHOR INFORMATION

José Luis Guzmán (joguzman@ual.es) earned a computer science engineering degree in 2002 and a Ph.D. in 2006, both from the University of Almería, Spain, where he is an assistant professor and a researcher in the Automatic Control, Electronics, and Robotics Group. Currently his interests center on control education and robust model predictive control techniques, with applications to interactive tools, virtual and remote labs, and agricultural processes. He can be contacted at the Universidad de Almería, Dpto. de Lenguajes y Computación, Ctra. Sacramento s/n, 04120, Almería, Spain.

Karl Johan Åström received his M.Sc. in engineering physics in 1957 and his Ph.D. in control and mathematics in 1960 from the Royal Institute of Technology in Stockholm, Sweden. After graduating, he worked for IBM Research for five years. In 1965, he became professor at Lund Institute of Technology/Lund University, where he founded the Department of Automatic Control. He is now emeritus at Lund University and a part-time visiting professor at the University of California at Santa Barbara. He is a Life Fellow of the IEEE. He has broad interests in control and its applications. He has received many awards, including the IEEE Medal of Honor and the Quazza medal from IFAC.

Sebastián Dormido holds a degree in physics from the Complutense University in Madrid, Spain (1968) and a Ph.D. from the University of the Basque Country, Spain (1971). In 1981 he was appointed professor of control engineering at the Universidad Nacional de Educación a Distancia. His activities include computer control of industrial processes, model-based predictive control, robust control, and model and simulation of continuous processes. He has authored and coauthored more than 150 technical papers in international journals and conferences. From 2002 to 2006 he was president of the Spanish Association of Automatic Control CEA-IFAC.

Tore Hägglund received his M.Sc. in engineering physics in 1978 and his Ph.D. in automatic control in 1984, both from Lund Institute of Technology, Sweden, where he is currently a professor. Between 1985 and 1989 he worked for Alfa Laval Automation (now ABB) on the development of industrial adaptive controllers. He is a coauthor of *Advanced PID Control* and *Process Control in Practice*. His current research interests are in the areas of process control, tuning and adaptation, and supervision and detection.

Manuel Berenguel is a professor at the University of Almería, Spain. He earned an industrial engineering degree and Ph.D. from the University of Seville, Spain, where he has been a researcher and associate professor for six years. His research interests are in predictive, adaptive, and robust control, with applications to solar energy systems, agriculture, and biotechnology. He has authored and coauthored more than 100 technical papers in international journals and conferences and is coauthor of the book *Advanced Control of Solar Plants* (Springer, 1997).

Yves Piguet is CEO of Calerga Sàrl, a company he cofounded in 2001 to develop and commercialize scientific software such as Sysquake, of which he is the main developer. He received a diploma in microtechnics in 1991 and a Ph.D. degree in control systems in 1997, both from EPFL, Lausanne, Switzerland. His research interests include robust control, real-time controllers and their use in mobile robotics, and interactive software and its applications.

Dynamics of Multibody Systems

by JENS WITTENBURG
Reviewed by Arend L. Schwab
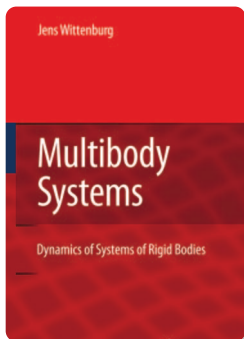
Dynamics of Multibody Systems, second edition
Springer-Verlag, 2008
ISBN 978-3540739135
US$99, XIII+223 pages.

This book is the second edition of the 1977 *Dynamics of Systems of Rigid Bodies* [1], which was written before the field of multibody system dynamics had adopted its current name. Indeed, the first edition was recognized as a forerunner or trendsetter in this field.

When I was performing my M.Sc. research on the dynamics of flexible mechanisms, around 1982, the original edition was an invaluable reference. I was particularly fond of the treatment of angular orientation of a rigid body as well as the use of Euler-Rodrigues parameters (quaternions). I clearly remember Prof. Wittenburg visiting the University of Delft at that time to give a mechanics colloquium on the dynamics of systems of rigid bodies with constraints. He demonstrated his formalism by means of simulation results from a Fortran program, the forerunner of the software tool Mesa Verde. For the 1980s, his examples were very challenging, for example, the dynamic analysis of an overconstrained 3D closed-loop kinematic chain consisting of six bodies and six joints; the dynamics of a three-legged swivel-wheel table (a nonholonomic system); and the impact analysis of a long chain struck by a point mass. All these examples could be found in the original book.

In the current second edition, chapters 1–4 are unchanged from the first edition apart from the addition of a short sections on quaternions and instantaneous screw axes. The last two major chapters, however, which cover general multibody systems and impact problems in multibody systems, include substantial revisions. Chapter 1 describes, in a rigorous way, the mathematical notation used throughout the book. In Chapter 2 rigid body kinematics is treated with a detailed description of the various angular orientation descriptions, including Euler angles, Bryan angles, and Euler-Rodrigues parameters. To make full use of Euler-Rodrigues parameters, a new section on quaternion algebra is added. Chapter 3, which is a short chapter on the basic principles of rigid body dynamics, closes with the principle of virtual power. In Chapter 4 some classical problems in rigid body mechanics are treated, including the torque-free rigid body, the symmetric heavy top, and the gyrostat.

Next, an extensive Chapter 5 treats general multibody dynamics for rigid bodies with constraints. In describing the system topology, the author makes use of graph theory. The equations of motion are derived by means of the principle of virtual power, and are expressed in terms of generalized independent coordinates where both tree and closed-loop structures are treated. Both holonomic constraints and nonholonomic constraints, which arise in idealized rolling contact, are treated. This chapter ends with a small section on vibration analysis of chains of bodies. In the last chapter (Chapter 6) impact problems in general multibody systems are treated.

DISCUSSION

This text is a classical and complete book on rigid multibody dynamics. Flexible bodies are not treated, but flexibility is taken into account in the form of massless springs, which suffices for many engineering applications. The book is precise in notation and clear in presentation. The use of graph theory for describing system topology is a bit cumbersome and distracts somewhat from the main goal, that is, dynamics of multibody systems. The reference list, of which 35% is not in the English language, is extensive and up-to-date. Surprisingly, however, some relevant references published after 1977 are not mentioned in the text. A particular example is the treatment of general impact in multibody systems in the 1996 book [2], which should have been explicitly cited in the last chapter on impact problems.

Without doubt, this classical textbook in multibody dynamics is an excellent reference, containing valuable material for graduate students and researchers. The chapters have an average of six example and problem sets, for which the solutions are provided. Aside from the lack of problems I believe that the book can easily be used as a text for a graduate course on the dynamics of rigid multibody systems.
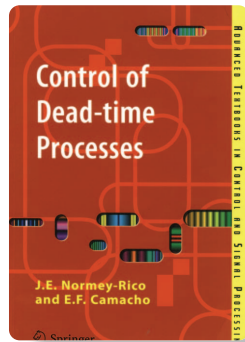
REFERENCES

[1] J. Wittenburg, *Dynamics of Systems of Rigid Bodies*. Stuttgart: Teubner, 1977.
[2] F. Pfeiffer and C. Glocker, *Multibody Dynamics with Unilateral Constraints*. New York: Wiley, 1996.

REVIEWER INFORMATION

Arend L. Schwab is an assistant professor in applied mechanics at Delft University of Technology in the Department of Mechanical Engineering. He is interested in multibody dynamics, in particular, contact phenomena such as collisions and rolling (nonholonomic constraints). His interests also include the dynamics of flexible multibody systems, finite element methods, legged locomotion, and bicycle dynamics. His degrees are from Dordrecht (B.Sc. 1979) and Delft (M.Sc. 1983, Ph.D. 2002).

Control of Dead-Time Processes

**BY J.E. NORMEY-RICO
AND E.F. CAMACHO
Reviewed by Keqin Gu**

Dead time, also known as a time delay, is present in many practical control systems. Dead time may be caused by the time needed to transport materials or to transmit or process information. In practice, dead time is also used to approximate finite-dimensional components such as a series of first-order systems. The continuous-time transfer function for systems with dead time is irrational, which presents significant challenges in analyzing and controlling such systems. Substantial effort has been devoted by the research community on such systems. See, for example, [1].

As a special case of systems with dead time, many practical industrial open-loop processes can be effectively described by a low-order transfer function in cascade with a dead time. Such a description seems to have its origin in the paper [2] (see also [3]) by Ziegler and Nichols, where a method for setting proportional intefgral derivative (PID) control parameters is also described. PID control remains the most popular control method in practice, and Ziegler and Nichols's tuning rules are presented in many introductory undergraduate control textbooks today. Not surprisingly, a substantial amount of research has been devoted to the topic of PID tuning over the years, as summarized in the book [4], as well as the recent special issue on PID control in this magazine [5].

In spite of its popularity, the performance of PID control has its limitations for systems with long dead times. Recent books [6] and [7] provided a method for determining the range of PID parameters such that the system is stabilized. The Smith predictor (SP) [8] promised to overcome the limitations associated with PID controllers for systems with long dead times and completely compensate for the effect of dead time for open-loop stable systems. This promise, however, has not been completely realized. While the logical simplicity of SP is appealing, its control structure is more complicated than PID control. Furthermore, the achievement of good control performance depends on a precise knowledge of the plant. In particular, SP can be very sensitive to the error of the dead time estimate, a subject of much research interest over the years. Efforts have also been made to extend the Smith predictor to modified and unified Smith predictors to handle open-loop unstable systems, as well as to enhance robustness against parameter errors [9].

A more modern approach to compensating for dead time is model predictive control (MPC). Due to its ability to handle multi-input, multi-output (MIMO) systems and input-output constraints, MPC has increasingly been adopted in practice.

GENERAL COMMENTS

Control of Dead-Time Processes presents the above three widely used control strategies in a textbook format, which until recently were available only through multiple sources such as [4] and [10]–[13] as well as articles in technical journals and conference proceedings. This treatment is a welcome addition to the literature, especially for those who are interested in learning the prevailing methods and practices of process control. The authors have made a commendable effort to keep the required background to a minimum. Indeed, anyone with a background in linear systems and discrete-time systems from a typical engineering curriculum should be able to follow the book.
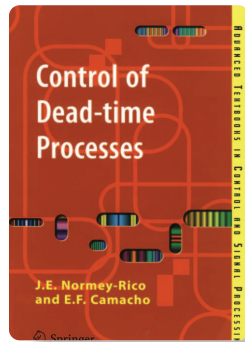
The book contains a generous number of practical examples. For most methods discussed, Matlab codes are also given. Heuristic reasoning and simulation illustrations are often used to motivate critical ideas. Many practical issues are also discussed. These features make the book especially appealing to those with an interest in immediate application.

The material is well organized and nicely presented. The editorial quality is high with very few mistakes for a first edition. My only complaint is the authors' tendency in a few places to use a sequence of two letters, such as na, to represent a single variable, which is inconsistent with the practice in most parts of the book, where the more traditional form of one letter with subscript, such as n_a, is used. Overall, this book is a pleasure to read for both beginners and experts. The book is

REVIEWER INFORMATION

Arend L. Schwab is an assistant professor in applied mechanics at Delft University of Technology in the Department of Mechanical Engineering. He is interested in multibody dynamics, in particular, contact phenomena such as collisions and rolling (nonholonomic constraints). His interests also include the dynamics of flexible multibody systems, finite element methods, legged locomotion, and bicycle dynamics. His degrees are from Dordrecht (B.Sc. 1979) and Delft (M.Sc. 1983, Ph.D. 2002).

Control of Dead-Time Processes

**BY J.E. NORMEY-RICO
AND E.F. CAMACHO
Reviewed by Keqin Gu**

Springer-Verlag
London, 2007
ISBN 978-1-84628-828-9,
US$99.00, 462 pages.

Dead time, also known as a time delay, is present in many practical control systems. Dead time may be caused by the time needed to transport materials or to transmit or process information. In practice, dead time is also used to approximate finite-dimensional components such as a series of first-order systems. The continuous-time transfer function for systems with dead time is irrational, which presents significant challenges in analyzing and controlling such systems. Substantial effort has been devoted by the research community on such systems. See, for example, [1].

As a special case of systems with dead time, many practical industrial open-loop processes can be effectively described by a low-order transfer function in cascade with a dead time. Such a description seems to have its origin in the paper [2] (see also [3]) by Ziegler and Nichols, where a method for setting proportional intefgral derivative (PID) control parameters is also described. PID control remains the most popular control method in practice, and Ziegler and Nichols's tuning rules are presented in many introductory undergraduate control textbooks today. Not surprisingly, a substantial amount of research has been devoted to the topic of PID tuning over the years, as summarized in the book [4], as well as the recent special issue on PID control in this magazine [5].

In spite of its popularity, the performance of PID control has its limitations for systems with long dead times. Recent books [6] and [7] provided a method for determining the range of PID parameters such that the system is stabilized. The Smith predictor (SP) [8] promised to overcome the limitations associated with PID controllers for systems with long dead times and completely compensate for the effect of dead time for open-loop stable systems. This promise, however, has not been completely realized. While the logical simplicity of SP is appealing, its control structure is more complicated than PID control. Furthermore, the achievement of good control performance depends on a precise knowledge of the plant. In particular, SP can be very sensitive to the error of the dead time estimate, a subject of much research interest over the years. Efforts have also been made to extend the Smith predictor to modified and unified Smith predictors to handle open-loop unstable systems, as well as to enhance robustness against parameter errors [9].

A more modern approach to compensating for dead time is model predictive control (MPC). Due to its ability to handle multi-input, multi-output (MIMO) systems and input-output constraints, MPC has increasingly been adopted in practice.

GENERAL COMMENTS

Control of Dead-Time Processes presents the above three widely used control strategies in a textbook format, which until recently were available only through multiple sources such as [4] and [10]–[13] as well as articles in technical journals and conference proceedings. This treatment is a welcome addition to the literature, especially for those who are interested in learning the prevailing methods and practices of process control. The authors have made a commendable effort to keep the required background to a minimum. Indeed, anyone with a background in linear systems and discrete-time systems from a typical engineering curriculum should be able to follow the book.

The book contains a generous number of practical examples. For most methods discussed, Matlab codes are also given. Heuristic reasoning and simulation illustrations are often used to motivate critical ideas. Many practical issues are also discussed. These features make the book especially appealing to those with an interest in immediate application.

The material is well organized and nicely presented. The editorial quality is high with very few mistakes for a first edition. My only complaint is the authors' tendency in a few places to use a sequence of two letters, such as na, to represent a single variable, which is inconsistent with the practice in most parts of the book, where the more traditional form of one letter with subscript, such as n_a, is used. Overall, this book is a pleasure to read for both beginners and experts. The book is

suitable as a textbook for a graduate level course or as a reference for practicing engineers. The first eight chapters can also be considered as a textbook for a final year undergraduate course.

It is also interesting to point out what this book is not about. This book is not about the fundamental properties of time-delay systems. It does not discuss fundamental issues such as initial conditions, the distribution of poles, definition of stability, and solution structures. The authors are satisfied with making one single citation [1] covering these topics. Nor is it a book on controller synthesis using a modern paradigm such as an H_∞ formulation; interested readers can find this approach in [9].

CONTENTS

After a short introduction in Chapter 1, Chapter 2 presents some examples of processes with dead times. Also discussed are some approximations of dead times, some of which serve to motivate the design of PID control in Chapter 4. Chapter 3 discusses identification of dead-time processes, including some simple methods, as well as the least squares method. Chapter 4 presents various design methods and configurations of PID control for several types of systems. A few practical case studies are also included.

Chapters 5–8 discuss dead-time compensators. Chapter 5 discusses properties of the SP, including modifications for unstable open-loop systems, especially integrative open-loop systems. Chapter 6 discusses modifications of SP for open-loop stable systems. Chapter 7 discusses the modified SP for open-loop unstable systems. Chapter 8 discusses discrete-time dead-time compensation (DTC).

Chapters 9 and 10 discuss model predictive control. Chapter 9 introduces two of the most common MPC control algorithms, dynamic matrix control (DMC) and generalized predictive control (GPC). The connection between MPC and DTC discussed in chapters 5–8 is established. Based on this connection, Chapter 10 discusses robust analysis and control design.

Chapters 11 and 12 deal with linear MIMO systems. Chapter 11 generalizes DTC to the MIMO case. After a brief introduction of the MIMO formulation, this chapter discusses single delay and multiple delay cases, and culminates with design and analysis of generalized multi-dead-time compensators.

Chapter 12 discusses MPC for MIMO systems. After a brief introduction of discrete-time MIMO formulation, this chapter presents the MIMO dead-time compensator-based generalized predictive controller (MIMO-DTC-GPC). A significant part of this chapter is devoted to the robustness analysis of MIMO-DTC-GPC. Implementation issues and case studies are also discussed.

Chapter 13 discusses generalizations of DTC and MPC to nonlinear systems and systems with constraints. The first part of the chapter discusses systems with constraints on inputs and other variables, while the later part describes nonlinear models and discusses the main issues that arise in extending DTC and MPC to such systems.

Chapter 14 discusses predictors for control purposes. It is shown that an "optimal predictor" in open loop may not be optimal in closed loop. Methods for integrating the design of the predictor and control are presented.

CONCLUSIONS

This well-written textbook presents PID control, the SP, and model predictive control in a single book. The book requires modest background to understand, and is suitable as a textbook for a final year undergraduate class or graduate class on process control. It is also valuable reference for practicing engineers.

REFERENCES

1] J.-P. Richard, "Time delay systems: an overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, 2003.
[2] J.G. Ziegler and N.B. Nichols, "Optimum settings for automatic controllers," *Trans. ASME*, vol. 64, pp. 759–768, Nov. 1942.
[3] J.G. Ziegler and N.B. Nichols, "Optimum settings for automatic controllers," *ASME J. Dyn. Syst. Meas. Contr.*, vol. 115, no. 2(B), pp. 220–222, 1993.
[4] K.J. Astrom and T. Hagglund, *PID Controllers: Theory, Design and Tuning*. Research Triangle Park, NC: Instrum. Soc. Amer., 1995.
[5] "PID 2006 Special Section," *IEEE Control Syst. Mag.*, vol. 26, no. 1, 2006.
[6] A. Datta, M.T. Ho, and S.P. Bhattacharyya, *Structure and Synthesis of PID Controllers*. New York: Springer-Verlag, 2000.
[7] G.J. Silva, A. Datta, and S.P. Bhattacharyya, *PID Controllers for Time-Delay Systems*. Cambridge, MA: Birkhaüser, 2005.
[8] O.J.M. Smith, "Closed control of loops with dead time," *Chem. Eng. Progress*, vol. 53, pp. 217–219, 1957.
[9] Q.-C. Zhong, *Robust Control of Time-Delay Systems*. New York: Springer-Verlag, 2006.
[10] Z.J. Palmor, "Time-delay compensation," in *The Control Handbook,* S. Levine, Ed. New York: CRC Press and IEEE Press, 1996, pp. 224–237.
[11] J.M. Maciejowski, *Predictive Control with Constraints*. Englewood Cliffs, NJ: Prentice-Hall, 2001.
[12] E.F. Camacho and C. Bordons, *Model Predictive Control*. New York: Springer-Verag, 2004.
[13] M. Morari and E. Zafiriou, *Robust Process Control*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

REVIEWER INFORMATION

Keqin Gu (kgu@siue.edu) is a professor and chair of the Department of Mechanical and Industrial Engineering, Southern Illinois University, Edwardsville. He received the B.S. and M.S. from Zhejiang University and the Ph.D. from Georgia Institute of Technology. His research interests include dynamics, control, and robotics, with emphasis on time-delay systems. He is a coauthor of the book *Stability of Time-Delay Systems*. He is currently an associate editor of *Automatica* and was an associate editor of *IEEE Transactions on Automatic Control*.
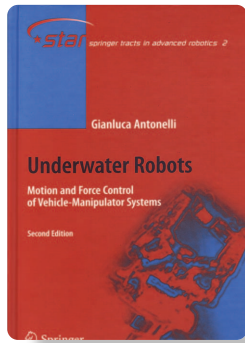
Underwater Robots: Motion and Force Control of Vehicle-Manipulator Systems

by GIANLUCA ANTONELLI

Reviewed by

Alexander Leonessa

Many aerospace engineers with an expertise in flight control system design, if tasked with designing a guidance system for an underwater vehicle, would dismiss the problem as easily solvable. To aerospace engineers, underwater vehicles are thought to be nothing more than powered blimps, which have been flying since 1852, when Henri Giffard built the first powered airship, a 143-ft long, cigar-shaped, gas-filled bag with a propeller that was powered by a 3-hp steam engine. If we can control intrinsically unstable airplanes at speeds several times that of sound, how difficult can it be to design a guidance system for an underwater vehicle that travels at less than 1/100 of the speed of sound? Of course, the fact that the density of water is about 800 times that of the air is irrelevant! As one of *those* aerospace engineers, after spending the last ten years *trying* to design that guidance system, I can finally recognize how "little I knew," and I thank Gianluca Antonelli for helping me understand the challenges that such a problem presents.

In this book, the author goes beyond the problem of controlling a single underwater vehicle by addressing the control of underwater vehicle/manipulator systems (UVMs). These systems have been used for many years in teleoperated versions to inspect and repair underwater cables and pipes, in search and rescue missions, for underwater archeology, and anything else that requires going underwater and grabbing something. A few years ago an operator of one of these teleoperated UVMs explained to me that it is very difficult to control the entire system simultaneously, so much so that, in practice, at first the manipulator is locked in place and the vehicle is moved until the end effector gets close to the desired position, then the manipulator is controlled toward the final target while the vehicle is maintained as steady as possible. By splitting the problem into two simpler problems, operators, who must be highly trained and talented, are able to complete the task. What makes this problem so difficult is the complexity of UVM dynamics, which are often redun-

dant and provide multiple possible solutions to the task and even more possible undesired outcomes. We also need to consider the difficulties related to limited communication with underwater systems, the hostility of the ocean environment, and the delays experienced in the control loops, just to mention a few. In an effort to overcome these difficulties, this book is aimed at control for *autonomous* UVMs. The author has done an excellent job addressing several of these challenges and providing possible solutions of increasing complexity as the book progresses.

In this second edition, the author has addressed many comments made by readers and reviewers by streamlining and improving the content. He has also added a few chapters addressing the state of the art and additional challenges, such as fault tolerance and collaborative control.

ABOUT THE BOOK

Our libraries are filled with excellent books discussing robot kinematics, dynamics, trajectory tracking, and with a good balance between theory and applications. However, the area of robot control is not dealt with as extensively; [1]–[3] provide noticeable exceptions. *Underwater Robots* contains similar topics to those covered in [1]–[3] but also addresses the additional challenges of a mobile platform (the underwater vehicle) and considers the difficulties and adversity related to the underwater environment.

The book begins by presenting a short discussion on the state of underwater vehicle technology. Topics such as sensors, actuators, localization, and control of underwater vehicles are briefly addressed with numerous references. A more formal definition of UVMs is also provided with a clear statement of this area as the core topic of the book.

Chapter 2 addresses modeling of UVMs. Representation of the rigid body kinematics is provided using both Euler angles and quaternions. The notation used is compact, and the various frames of reference are clearly identified, which makes the notation clear and facilitates understanding of the following chapters. The treatment of rigid body dynamics starts from first principles and does not assume much previous knowledge on the topic other than Newton's law, which makes this chapter particularly attractive for classroom use. Hydrodynamic effects are briefly discussed, including added mass, damping, currents, and buoyancy. At this point the resulting model is similar to that found in [4]. However, the kinematics and dynamics of the manipulator are then introduced, including the coupled dynamics of the vehicle/manipulator as well as additional phenomena, such as contact with the environment. The overall presentation is thus much more general than the treatment in [4].

In Chapter 3 a survey of existing control algorithms for autonomous underwater vehicles (AUVs) is presented. Various frames of reference, model- and nonmodel-based, full- and reduced-order algorithms are considered as well as compensation of ocean currents. All of the results are

rigorously proven using a Lyapunov-function approach. What makes this chapter especially interesting is a qualitative comparison of these controllers. The code used for the simulation is made available on a website.

Chapter 4, which is one of the new chapters of this second edition, addresses fault detection and tolerance strategies for underwater vehicles. Both sensor and actuator failures are discussed, and a survey of various schemes is presented with a plethora of references. The final section contains some experimental results, which nicely prepares the reader for the next chapter.

In Chapter 5 the author provides some experimental results on both control and fault tolerance to thruster faults. The experiments were conducted at the University of Hawaii using the Omni-Directional Intelligent Navigator (ODIN), which unfortunately does not have a manipulator. The practical aspects of the implementation are a welcome addition to this chapter.

Chapters 6–8 focus on the core topics of this book and provide the kinematics, dynamics, and interaction control for UVMs, respectively. The three chapters are well laid out, and it is easy to see conceptually where the author is going to take you. From a technical point of view the derivation can be quite involved at times, especially considering the theoretical rigor of these chapters, but the notation introduced in the preceding chapters helps to make the presentation easy to follow. In particular, kinematic control is presented to allow real-time trajectory planning and account for redundancy. Several kinds of Jacobian pseudoinverse approaches are introduced as well as drag minimization and task-priority algorithms. Several case studies are provided to demonstrate the benefits of each algorithm. Dynamic control is then discussed. Many control algorithms are presented, such as feedforward decoupling, feedback linearization, sliding mode, and adaptive control. It is worth mentioning that classical control strategies developed for industrial robotics cannot be directly used on UVMs because of several issues, which include reduced knowledge of hydrodynamics effects, poor thruster performance, and dynamic coupling between vehicle and manipulator. These challenges require the introduction of novel control strategies that address tracking performance, which must remain simple enough to be implementable using the limited memory and computational power generally available onboard these vehicles. In particular, the author observes that classical adaptive control strategies applied to the UVMs as a whole generate very high dimensional problems, which require unreasonable computational loads to be solved. This issue is addressed by introducing a virtual decomposition approach, which exploits the serial-chain structure of the UVMS by decomposing the overall motion control problem into a set of simpler problems, addressing the

motion of each manipulator link and the vehicle. Finally, interaction with the environment is discussed, which is a necessary step for controlling a manipulator. Impedance, force, and external force control algorithms are discussed, and robustness considerations, implementation issues, and simulations are provided for each of them.

Chapter 9 is the last of the additional chapters of the revised edition. At first, I was skeptical about having a short chapter on coordinated control of a platoon of AUVs in a book whose focus is on UVMs. However, after reading the chapter, I understood the intent of the author to finish his book with a chapter describing future challenges and endeavors. This chapter provides a good survey of *practical* implementations of coordinated control rather than complex theoretical results. Numerous references are provided for a starting point in a field that is characterized by exponential growth in interest by numerous research groups.

CONCLUSIONS
Underwater Robots does an excellent job presenting many of the issues related to the modeling and control of UVMs. The text provides a good balance between basic results that can be used in teaching an advanced class on this topic as well as more advanced results targeting experimental researchers looking for a particular control law to implement on their system. The number of references is impressive, and most of the work in the field that any researcher would need to consult for a deeper understanding of the state of the art is included. Finally, although the book is focused on systems with manipulators, it is also a good source for readers interested in the general field of autonomous underwater vehicles.

REFERENCES
[1] H. Asada and J.-J.E. Slotine, *Robot Analysis and Control*. New York: Wiley, 1986.
[2] M.W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Dynamics and Control*. New York: Wiley, 2005.
[3] L. Sciavicco and B. Siciliano, *Modelling and Control or Robot Manipulators*. New York: Springer-Verag, 2000.
[4] T.I. Fossen, *Guidance and Control of Ocean Vehicles*. New York: Wiley, 1994.

REVIEWER INFORMATION
Alexander Leonessa is an assistant professor in the Mechanical Engineering Department at Virginia Tech. He received the Laurea from the University of Rome "La Sapienza" and the M.S. and Ph.D. from Georgia Tech. His areas of expertise include control theory, robotics, and mechatronics, with applications to propulsion systems, autonomous vehicles, attitude stability and control, robot control, human-robot interaction, and functional electrical stimulation.
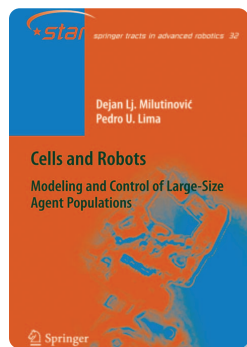
Cells and Robots: Modeling and Control of Large-Size Agent Populations

by DEJAN LJ. MILUTINOVIĆ
and PEDRO LIMA
Reviewed by Nikolaus Correll

The relation between robots and cells suggested by the book title might be surprising since it is probably not apparent at first sight. Thinking about this more closely, however, the following facts come to mind. Single-cell organisms such as bacteria explore their environment by gradient descent toward nutrition sources or magnetic fields and communicate with each other through inter-cellular channels. These mechanisms, as well as collective natural phenomena at the nanoscale such as the immune system, have already inspired the design of algorithms and subsystems for robot swarms. Where these capabilities might also lead us is well summarized by the title of H.C. Berg's keynote talk titled "Motile Behavior of E. Coli, a Remarkable Robot" at the 2008 Robotics: Science and Systems Conference [1]. Moreover, the emerging field of synthetic biology aims at making the design of cells with specific properties an engineering discipline, and it is thus likely that future nanorobotic swarms will rely on a significant biological component for sensing and actuation.

For these reasons, models developed for collective cellular systems might be potentially applicable to robotic swarms and vice versa. In particular, a mathematical model of such systems might help us to formally understand the relation between individual and collective dynamics, which is a challenging question among multiple disciplines, whether considering the immune system or robotic swarms. Milutinović and Lima's book is a step forward toward the solution of this quest. The book introduces a hybrid [2] dynamical modeling framework for modeling both the discrete population dynamics as well as the distribution of the swarm in a continuous state space. This approach differs from previous work, which aims at either modeling the fraction of swarm members in a given discrete behavioral state (for example, [3] for a difference equation model for a swarm-robotics case study) or modeling the distribution of the swarm in a continuous state space (for example, [4] for differential equation models commonly used in virology and

immunology). The authors assume that individual members of the swarm switch between discrete states probabilistically. They further assume that a vector field is associated with each discrete state and that this vector field is responsible for the temporal evolution of the distribution in the continuous space. The authors then develop a system of partial differential equations (PDEs) that describe the change of the continuous distribution of each discrete state as a function of both the state transition probabilities between discrete states and the vector fields that affect the continuous state distributions. This approach is illustrated using two case studies, the distribution of T-cell expression levels in the immune system and the spatial distribution of a swarm of robots. Here, both T-cells and robots can be in one of multiple discrete behavioral states, whereas the continuous state distribution describes the expression level of the T-cell population or the location of the robots.

CONTENTS

After introducing the reader briefly to the analogy between an individual robot and a cell in terms of sensors and actuators in Chapter 1 and the immune system and T-cell receptor dynamics in Chapter 2, Chapter 3 introduces the hybrid automata approach that is used for modeling the individual swarm member throughout the book. Chapter 4 then describes the relation between the hybrid automaton that describes the individual dynamics and the macroscopic dynamics. This relation is captured by the PDE model described above. The developed method is then applied to the T-cell expression case study in Chapter 5. Noteworthy contributions of this chapter are the validation of the PDE approach using experimental data as well as the comparison of the PDE model that explicitly describes the distribution of T-cell expression dynamics with commonly used ordinary differential equation (ODE) models. ODE models are limited to describing average quantities, whereas the PDE system also describes the distributions of these quantities. Chapter 6 then explores populations with heterogeneous parameters and extends the modeling framework by explicitly modeling the resulting parameter uncertainty in the PDE system.

The primary goal of modeling the T-cell population in the immune system is to achieve better systems understanding and prediction of laboratory observations. For fully engineered systems, such as a robot swarm, models can serve an additional purpose, namely, they can be used for the design of the individual agent. Using methods from optimal control (in particular the minimum principle for PDEs), Chapter 7 shows how the *individual* state transitions need to be tuned in an open-loop control scheme to achieve a desired distribution in the continuous state space. Conclusions given in Chapter 8 are followed by appendices A–C, which detail experimental and analysis methods relevant to T-

cell experiments. Finally, Appendix D reviews the minimum principle for PDEs.

AUDIENCE

This book addresses the advanced researcher or graduate student who is interested in a probabilistic perspective on modeling of large-scale distributed systems. Appreciating the book to its full extent, however, requires a solid background in PDEs and optimal control theory. Similarly, following the reasoning of the author is often difficult and depends heavily on the reader's background, for example, when biological or robotic knowledge is required to follow the intuition behind the presented models. Although Chapter 1 and Chapter 2 provide a brief overview of robotic sensors and actuators as well as the relevant aspects of T-cell dynamics, respectively, I personally needed the intuition provided by the robotics example in Chapter 7 to fully understand the PDE modeling formalism, which in turn helped me to understand the immune system case study. Fortunately, each chapter in the book is self-contained, which caters to the reader who wants to read chapters selectively.

CONCLUSIONS

One might argue that the case studies brought forward by the authors are very specific and that the robotic case study in particular might be of limited application and lacks validation at a lower modeling abstraction level. Nevertheless, this book elaborates on two important points. First, for the biologist and the robotics community, the authors introduce an extremely compact model that describes the evolution of the probabilistic state distribution o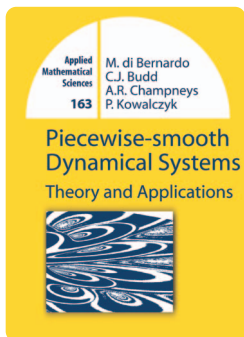f a multiagent system with discrete and continuous states. Second, by controlling average quantities and their distributions rather than deterministically controlling the individual agent, the authors advance the state of the art in modeling large-scale distributed systems whose agents are of limited capabilities and subject to sensor and actuator noise. In summary, I recommend this book to anyone who is interested in a probabilistic perspective on modeling large-scale distributed systems–an area that definitely deserves more attention.

REFERENCES

[[1] "Robotics science and systems" [Online]. Available: http://roboticsconference.org/
[2] A.J. van der Schaft and J.M. Schumacher, *An Introduction to Hybrid Dynamical Systems.* New York: Springer-Verlag, 2000.
[3] A. Martinoli, K. Easton, and W. Agassounon, "Modeling swarm-robotic systems: A case study in collaborative distributed manipulation," *Int. J. Robotics Res.*, vol. 23, pp. 415–436, 2004.
[4] M.A. Nowak and R.M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology.* London, U.K.: Oxford Univ. Press, 2000.

REVIEWER INFORMATION

Nikolaus Correll is a postdoctoral fellow at the Distributed Robotics Laboratory, Computer Science and Artificial Intelligence Laboratory, MIT. He earned an M.S. in electrical engineering from the Swiss Federal Institute of Technology (ETH) in Zurich in 2003 and a Ph.D. in computer science from the Ecole Polytechnique Federale Lausanne (EPFL) in 2007. His research interests include large-scale distributed robotic systems, mixed animal-robot societies, and monitoring of collective systems. He is technical program cochair for the nanorobotics track at Nano-Nets 2008 in Boston, and serves as reviewer for several international robotics journals and conference proceedings.

Piecewise-Smooth Dynamical Systems: Theory and Applications

by MARIO DI BERNARDO, CHRISTOPHER J. BUDD, ALAN R. CHAMPNEYS, and PIOTR KOWALCZYK
Reviewed by Bernard Brogliato

Springer-Verlag, 2008
ISBN 978-1-84628-039-9
US$99.00.

This book deals with the analysis of bifurcations and chaos in nonsmooth (called piecewise smooth by the authors) dynamical systems. Despite the fact that nonsmooth systems have become a major topic in the control community [6], there remains no precise definition of what systems or more accurately, of what models are nonsmooth. Nonsmooth systems are usually represented by mathematical formalisms, such as switching systems, piecewise *something* systems, differential inclusions (there are various types of these depending on the properties of the set-valued right-hand-side), impulsive ordinary differential equations (ODEs), evolution variational inequalities, projected dynamical systems, complementarity dynamical systems (there are also various types of these), hybrid systems, and so on. My point here is that it soon becomes quite difficult to know precisely what a nonsmooth dynamical system is and how to classify them because nonsmooth systems are much like nonlinear systems relative to linear systems. That is, nonsmooth systems include everything that is not smooth, which yields a set with a lot of elements. Nevertheless, most nonsmooth formalisms share the property that they are not a simple extension of smooth formalisms developed for models whose right-hand-side possesses

cell experiments. Finally, Appendix D reviews the minimum principle for PDEs.

AUDIENCE

This book addresses the advanced researcher or graduate student who is interested in a probabilistic perspective on modeling of large-scale distributed systems. Appreciating the book to its full extent, however, requires a solid background in PDEs and optimal control theory. Similarly, following the reasoning of the author is often difficult and depends heavily on the reader's background, for example, when biological or robotic knowledge is required to follow the intuition behind the presented models. Although Chapter 1 and Chapter 2 provide a brief overview of robotic sensors and actuators as well as the relevant aspects of T-cell dynamics, respectively, I personally needed the intuition provided by the robotics example in Chapter 7 to fully understand the PDE modeling formalism, which in turn helped me to understand the immune system case study. Fortunately, each chapter in the book is self-contained, which caters to the reader who wants to read chapters selectively.

CONCLUSIONS

One might argue that the case studies brought forward by the authors are very specific and that the robotic case study in particular might be of limited application and lacks validation at a lower modeling abstraction level. Nevertheless, this book elaborates on two important points. First, for the biologist and the robotics community, the authors introduce an extremely compact model that describes the evolution of the probabilistic state distribution of a multiagent system with discrete and continuous states. Second, by controlling average quantities and their distributions rather than deterministically controlling the individual agent, the authors advance the state of the art in modeling large-scale distributed systems whose agents are of limited capabilities and subject to sensor and actuator noise. In summary, I recommend this book to anyone who is interested in a probabilistic perspective on modeling large-scale distributed systems–an area that definitely deserves more attention.

REFERENCES

[[1] "Robotics science and systems" [Online]. Available: http://roboticsconference.org/

[2] A.J. van der Schaft and J.M. Schumacher, *An Introduction to Hybrid Dynamical Systems*. New York: Springer-Verlag, 2000.

[3] A. Martinoli, K. Easton, and W. Agassounon, "Modeling swarm-robotic systems: A case study in collaborative distributed manipulation," *Int. J. Robotics Res.*, vol. 23, pp. 415–436, 2004.

[4] M.A. Nowak and R.M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology*. London, U.K.: Oxford Univ. Press, 2000.

REVIEWER INFORMATION

Nikolaus Correll is a postdoctoral fellow at the Distributed Robotics Laboratory, Computer Science and Artificial Intelligence Laboratory, MIT. He earned an M.S. in electrical engineering from the Swiss Federal Institute of Technology (ETH) in Zurich in 2003 and a Ph.D. in computer science from the Ecole Polytechnique Federale Lausanne (EPFL) in 2007. His research interests include large-scale distributed robotic systems, mixed animal-robot societies, and monitoring of collective systems. He is technical program cochair for the nanorobotics track at Nano-Nets 2008 in Boston, and serves as reviewer for several international robotics journals and conference proceedings.

Piecewise-Smooth Dynamical Systems: Theory and Applications

by MARIO DI BERNARDO,
CHRISTOPHER J. BUDD,
ALAN R. CHAMPNEYS, and
PIOTR KOWALCZYK
Reviewed by Bernard Brogliato

Springer-Verlag, 2008
ISBN 978-1-84628-039-9
US$99.00.

This book deals with the analysis of bifurcations and chaos in nonsmooth (called piecewise smooth by the authors) dynamical systems. Despite the fact that nonsmooth systems have become a major topic in the control community [6], there remains no precise definition of what systems or more accurately, of what models are nonsmooth. Nonsmooth systems are usually represented by mathematical formalisms, such as switching systems, piecewise *something* systems, differential inclusions (there are various types of these depending on the properties of the set-valued right-hand-side), impulsive ordinary differential equations (ODEs), evolution variational inequalities, projected dynamical systems, complementarity dynamical systems (there are also various types of these), hybrid systems, and so on. My point here is that it soon becomes quite difficult to know precisely what a nonsmooth dynamical system is and how to classify them because nonsmooth systems are much like nonlinear systems relative to linear systems. That is, nonsmooth systems include everything that is not smooth, which yields a set with a lot of elements. Nevertheless, most nonsmooth formalisms share the property that they are not a simple extension of smooth formalisms developed for models whose right-hand-side possesses

strong differentiability properties. This observation is true for control [6], for numerical methods [2], and also for dynamical analysis, as this book shows. Roughly speaking, what happens is that all the tools that are based on linearization, such as eigenvalues, eigenspaces, continuous dependence of eigenvalues on the parameter, center manifold reduction, and Taylor series expansion, no longer work for nonsmooth systems.

For a control scientist or an applied mathematician, a reasonable path is to rely on the applications and modeling work done in fields such as mechanics, electromagnetism, and biology, where nonsmooth models have been derived. For instance, mechanical systems subject to nonsmooth contact laws (impacts, or Coulomb friction are the most common), switched electrical circuits, and Filippov systems, which are familiar to systems and control researchers. Once an application has been targeted, a specific model can be chosen for control or analysis purposes. This approach is chosen in this book, which deals with bifurcation and chaos analysis for several classes of maps and systems, namely, C^0, discontinuous, square-root, and C^k maps, systems with C^0 nondifferentiable vector fields, Filippov systems, and vibro-impact systems. Strangely, the word "bifurcations" does not appear in the title of the book, despite the fact that this is the main topic being addressed. The general goal of this monograph is to present the peculiarities of bifurcations and chaos in such systems, in a rather detailed way, where many examples with detailed calculations, comments, and numerical results illustrate the theoretical results. Most of the theorems are stated without proof, but are illustrated through worked examples. It is clear that the authors have made a particular pedagogical effort in writing this book. The central topic, although not about feedback and control, is connected to control, and throughout the book the reader can find many notions that are widely used in control analysis, such as stability, canonical forms, relative degree, relay systems, Filippov systems, sliding modes, and Poincaré maps.

CONTENTS

The first chapter introduces the main notions that are used in the subsequent chapters, through the analysis of several examples, including impact oscillators, relay control, dry friction, dc-dc converter, and typical maps (square root and piecewise linear). Impact and stroboscopic Poincaré maps, grazing bifurcations, periodic orbits, numerical methods, limit sets, and penalized models are presented as well as experimental methods that are used to study and validate bifurcation diagrams. Readers who know the basics of bifurcation and chaos theory for smooth systems will certainly appreciate this chapter. Those who do not may first jump to the second chapter and then return to Chapter 1. Many fundamental tools and results concerning smooth systems are recalled in Section 2.1, which also gives insight into why bifurcation theory for smooth systems does not extend easily to nonsmooth ones. In particular, it is shown that a key result

obtained from the implicit function theorem and used in bifurcation theory for smooth systems no longer works (Theorem 2.4). Sections 2.2 and 2.3 review various mathematical formalisms for nonsmooth systems and maps. Sections 2.4 and 2.5 are dedicated to nonsmooth systems. The definition of a bifurcation, especially a discontinuity-induced bifurcation (DIB), is given. DIBs form a family of bifurcations that are typical to nonsmooth systems and that do not exist in smooth systems. Chapter 2 ends with a section on numerical methods that are used to integrate the systems of interest. Nonsmooth systems must be time discretized and integrated with great care [2], and most of the available software packages used in systems and control are not quite adequate.

I would make two main reproaches to chapters 1 and 2. First the difference between smooth and nonsmooth (piecewise smooth in the book's terminology) systems could be clearer. Definition 2.2 introduces smooth systems of index r, and Definition 2.20 concerns piecewise-smooth systems. It is not clear, however, whether the intersection between these is empty or not. A hint is given in a remark (page 50), where one understands that smooth means with a vector field that is C^k with $k > 0$, while the rest of the book confirms that piecewise smooth concerns vector fields that are C^0 or less regular. Second, a definition of bifurcations (the DIBs) in nonsmooth systems is given (Definition 2.32) based on the property of piecewise structural stability. Other researchers have argued (for example, [3]) that this definition is not the only possible one and may sometimes not be suitable. Some words on this point would have been welcome. Another point that seems surprising is the absence of Lyapunov exponents, which seem to be a central tool widely used by other authors [4], [5]. This is certainly a deliberate choice, which can be understood since focusing "only" on DIBs already fills almost the whole book. In conclusion, despite a few details that many readers will correct themselves, chapters 1 and 2 nicely introduce and motivate the topic of the book.

The remainder of the book (chapters 3–9) is dedicated to the very detailed study of DIBs in nonsmooth maps and nonsmooth systems. These chapters contain an enormous amount of information on DIBs, which clearly make this book a unique contribution to the field of bifurcations and chaos in nonsmooth systems. I confess I did not have the time to read it all with sufficient care to provide an in-depth review. I will therefore focus on a few parts only. However, since the chapters are presented in the same way, with theoretical aspects soon followed by simple examples and numerical results with discussions, I believe that observations that hold for these parts are representative of the remaining parts.

Chapter 6 deals with limit-cycle bifurcations in vibro-impact systems, which are simple mechanical systems that undergo a succession of free-motion trajectories and impacts and are as such highly nonsmooth and nonlinear; these systems are sometimes called flows with collisions [6]. The chapter starts with several examples of impacting systems, along with an introduction to suitable Poincaré

maps. The main point of this chapter is what happens close to grazing orbits, which are introduced in Chapter 1. Grazing trajectories touch the boundary, but with a zero normal velocity. It happens that the Poincaré maps that are associated with such grazing trajectories are piecewise smooth, linear on one side of the switching surface and of the square-root type on the other side. The square root function has an infinite gradient at zero, resulting in a stretching of the phase space in the zero limit. Intuitively, this property means that the dynamics of a system that evolves on a grazing trajectory (that is a trajectory that touches the boundary of the switching surface at one point, with a zero normal velocity) may vary in a very abrupt way when a parameter is changed, and the system undergoes a (strong) bifurcation, possibly chaos. The construction of the square-root maps is explained in detail in Section 6.2. Perhaps the presentation of this part could have been simplified by focusing directly on mechanical systems and not on an extension of them. Generally speaking, embedding nonsmooth mechanical systems in a hybrid framework is, in my humble opinion, quite useless. The chapter continues with the study of grazing bifurcations for periodic trajectories and the derivation of Poincaré maps.

The chapter ends with the influence of chattering, that is, accumulations of impacts, on the bifurcation process, as well as a short section on multiple impacts, that is, several collisions occurring at the same time in a system. Here it is not clear to me how accumulations are managed numerically, since it seems that event-driven methods [2] are used to get the bifurcation diagrams. This technique must be limited to simple cases where one knows what happens after the accumulation point. Also, since chattering is in a sense a plastic impact (zero restitution coefficient $r = 0$) obtained after an infinite number of impacts, I was wondering whether the case $r = 0$ would be equivalent to the case of chattering. Chapter 8 deals with sliding mode systems. A typical example is relay systems, which have been extensively studied in the control community from the point of view of existence of limit cycles.

I will end this part of the review with a few comments on Chapter 9. Section 9.1 discusses the possible discrepancies that arise between numerical and experimental results in simple impacting systems and provides explanations for these discrepancies, such as neglected dynamics and parameter uncertainties. Sections 9.3 and 9.4 present two applications and numerical calculations, namely, rattling gears and a hydraulic damper. The objective of Section 9.3 is to show that two different models (rigid body and restitution, and compliant contact) provide the same results. What is missing here is a conclusion on what model is preferable. Certainly, the numerical integration as the stiffness becomes large can be a criterion, because one would like to avoid integrating stiff ODEs and prefer specific methods for rigid-body

systems [2]. The chapter ends with examples of two-parameter sliding bifurcations in one degree-of-freedom mechanical systems with Coulomb friction, in contrast to the rest of the book, which is dedicated to one-parameter bifurcations. Chapter 9 looks like Chapter 1 in its construction and nicely closes the book.

CONCLUSIONS

This book is undoubtedly a strong contribution to the field of bifurcation and chaos analysis and more generally to the field of nonsmooth dynamical systems analysis. The authors have made a remarkable effort in mixing intricate technical developments with numerous examples, numerical results, and experimental results. It is obvious that the book is primarily intended for bifurcation and chaos specialists, for whom it will serve as a reference. Most systems and control researchers will certainly find parts of it a bit hard to read but once again the detailed examples help a lot. For beginners in the field, first reading a basic textbook on bifurcations and chaos (for instance, [7]) will certainly be quite helpful. Also, it is clear that control researchers would have appreciated a section on the control applications of bifurcations and chaos (the one that comes to my mind is the OGY method to stabilize chaotic systems on a periodic orbit, but there are more), for instance as a section of Chapter 9. The authors cannot be blamed for this omission, since feedback stabilization was not at all the primary objective of the book. Finally, the presentation, including both text and figures, is of high quality, and I found very few typos.

REFERENCES

[1] J. Cortes, "Discontinuous dynamical systems. A tutorial on solutions, nonsmooth analysis, and stability," *IEEE Control Syst. Mag.*, vol. 28, pp. 36-73, June 2008.

[2] V. Acary and B. Brogliato, *Numerical Methods for Nonsmooth Dynamical Systems.* New York: Springer-Verlag, 2008.

[3] R.I. Leine and H. Nijmeijer, *Dynamics and Bifurcations in Non-smooth Mechanical Systems*, vol. 18. New York: Springer-Verlag, 2004.

[4] J. Awrejcewicz and C.-H. Lamarque, *Bifurcation and Chaos in Nonsmooth Mechanical Systems.* Singapore: World Scientific, 2003.

[5] S.L.T. de Souza and I.L. Caldas, "Calculation of Lyapunov exponents in systems with impacts," *Chaos Solitons Fractals*, vol. 19, no. 3, pp. 569–579, 2004.

[6] B. Brogliato, *Nonsmooth Mechanics*, 2nd ed. New York: Springer-Verag, 1999.

[7] J.K. Hale and H. Kocak, *Dynamics and Bifurcations*, 3rd ed. (Texts in Applied Mathematics vol. 3). New York: Springer-Verag, 1996.

REVIEWER INFORMATION

Bernard Brogliato received the B.S. in mechanical engineering from the Ecole Normale Supérieure de Cachan (Paris) in 1987. He received the Ph.D. in automatic control from the National Polytechnic Institute of Grenoble (INPG) in 1991 and his Habilitation in 1995. He works at INRIA (the French National Institute for Research in Computer Science and Control) in Grenoble. His research interests are in nonsmooth dynamical systems modeling, analysis, and control, and dissipative systems.

The HYCON-EECI Graduate School on Control 2008

The second session of the HYbrid CONtrol-European Embedded Control Institute (HYCON-EECI) Graduate School met from February to April 2008. Graduate students were lectured by internationally renowned experts at Supelec, Gif sur Yvette, France. The number of participants increased from 111 in 2007 to 167 in 2008. Ten modules were offered this year, each providing over 20 hours of teaching.

Organized within the framework of the Network of Excellence, HYCON, and the European Embedded Control Institute (EECI), the graduate school follows the pathway of the FAP (Formation en Automatique de Paris) courses, which were launched by F. Lamnabhi-Lagarrigue (L2S, France) and coordinated by A. Loria (L2S, France) from 2004 to 2006. The objective of the graduate school is to provide Ph.D. students with a series of one-week lectures from experts in the field of control.

While offering an opportunity for participants to deepen their knowledge on a specific subject, the HYCON-EECI Graduate School on Control provides an opportunity for the participants to meet colleagues sharing similar interests and to validate modules now required in most European Ph.D. programs. The format of the courses allows for intensive study of a particular topic in only one week. In 2007 and 2008 the program was organized by J. Daafouz (CRAN, France). This year the program covered a wide scope of control theory including hybrid dynamics, linear matrix inequalities, optimal control, and geometric approaches.

Participants of the "LMIs in Control" module.
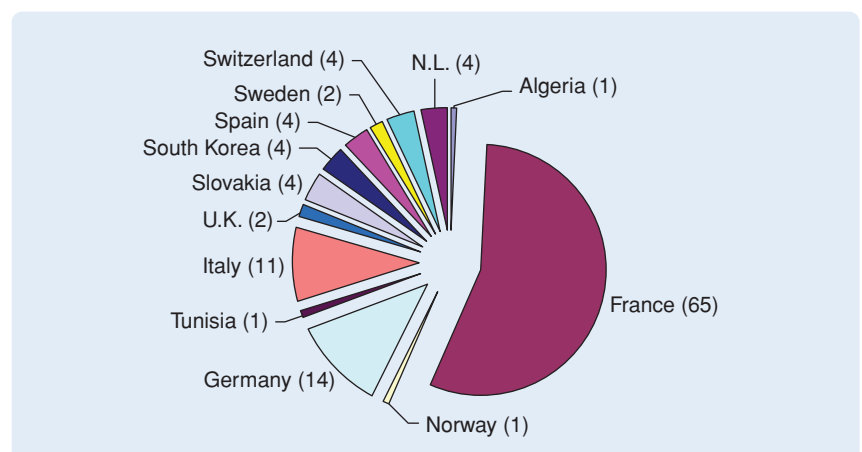
2008 HYCON participants by nationality.

Hybrid dynamical systems were the core of three different modules. E. De Santis (University of L'Aquila, Italy) focused on observability and observer synthesis. F. Wirth (University of Würzburg, Germany) presented sufficient Lyapunov conditions and converse results for the stability of hybrid systems, both in the linear and nonlinear cases. C. Prieur (LAAS, France) and A.R. Teel (UCSB, USA) presented a module on stability analysis and robust stabilization of hybrid systems.

Another important focus area was optimization and optimal control. Algorithms for local and global optimization, including an introduction to nonconvex optimization, were presented by F. Schoen (University of Firenze, Italy). Optimal control design was addressed by F. Clarke (University of Lyon 1, France) in the context of nonsmooth analysis and discontinuous feedback control of nonlinear plants.

Linear matrix inequality (LMI) techniques also constituted an important part of the courses taught this year. D. Henrion (LAAS, France) presented mathematical fundamentals for LMIs, together with optimization and robust stabilization methods. These subjects were further addressed by C. Scherer (Delft UT, the Netherlands) and S. Weiland (Eindoven UT, the Netherlands), who presented a course titled "LMIs in Control."

Additional modules included R. Banavar's (IIT Bombay, India) presentation of a module devoted to robotics, specifically, geometric approaches to control design for robots with nonholonomic constraints or underactuation. Finally, a module copresented by A. Bayen (Berkeley, USA) and I. Mitchell (University of British Columbia) discussed techniques for the safety verification of nonlinear systems and the synthesis of guaranteed safe control.

The next session of the HYCON-EECI Graduate School on Control will take place in the spring of 2009 at Supelec, Gif sur Yvette, France. Information on the graduate school and other activities of the EECI is available at www.eeci-institute.eu.

Antoine Chaillet
EECI Communication

Michael J. Rabins (1932–2008)

The international and ASME dynamic systems and control community has lost a truly inspiring and dedicated leader with the death of Michael Jerome Rabins. The qualities of his life can be characterized through his sincerity, loyalty, earnestness, enthusiasm, and friendship. He eagerly accepted responsibility, never refused to take on a hard job if it needed to be done, and was willing to make hard decisions. At the time of his death on September 30, 2007, Mike was an emeritus professor of the Department of Mechanical Engineering at Texas A&M University. He died of complications from leukemia.

Mike leaves a powerful legacy in the dynamic systems and control division of the ASME. His involvement began in 1963 by serving as the chair of publicity and the newsletter editor in the automatic control division. He served as secretary from 1968 to 1970. He was the founding editor of the ASME quarterly journal *Dynamic Systems, Measurement and Control* from 1970 to 1974. From 1972 to 1997 Mike was a member of the Automatic Control Division Executive Committee, becoming the chair from 1975 to 1976. From 1976 to 1981 he served as a member-at-large on the ASME Policy Board on Communications and from 1981 to 1984 was vice-president for communications. From 1983 to 1985 he was a committee member of the ASME board on research. In 1985 he was appointed to the new ASME board on research and technology development and served as chair of the Technology Opportunities Committee from 1985 to 1987. Additional-

ly, Mike served on the ASME Board of Governors in 1989–1990, was chair of the task force on fellow election procedure revision in 1990, and a member of the committee in 1991–1992, on the ASME Fellow Program Review Committee from 1991 to 1994, chair of the DSCD education panel from 1990 to 1993, and guest editor of the "50th Anniversary Special Issue of the *Journal of Dynamic Systems, Measurement and Control*."

Mike influenced numerous members in the dynamic systems and control community, many of whom learned by watching and listening to him at DSC Executive Committee meetings in which he always projected warmth, inclusion, openness of opinions, and integrity. Mike's personality was irresistible, comforting, reassuring, and engaging. He continued to be one of the few "elder" statesmen who actively participated in DSCD. Even after his retirement he served the division with distinction as a member of the DSCD Award Committee from 1997 to 2000.

Mike was chosen as a member of the Automatic Control Council/Exchange team to visit the U.S.S.R. in 1978. He served as the vice president of the American Automatic Control Council (AACC) from 1981 to 1983 and as president from 1983 to 1985. The last Joint Automatic Control Conference (JACC) was held in 1981 at the University of Virginia with ASME as the host society. In June 1982, the American Control Conference (ACC) replaced the JACC as the annual jointly sponsored national control conference. The ACC is organized under the direction of a steering committee, appointed directly by the AACC. Mike was the general chair of the first ACC in 1982, which was held in Arlington, Virginia.

Mike Rabins was born on February 24, 1932 in New York City. He attended the Bronx High School of Science, graduating in 1949. He then attended MIT where he received the B.S. in mechanical engineering in 1953. In 1954 he received the M.S. in mechanical engineering from the Carnegie Institute of Technology and finished the Ph.D. at the University of Wisconsin, Madison, in 1959. He served as an assistant professor at the University of Wisconsin from 1959 to 1960. He joined the faculty at New York University in 1960, where he was an assistant professor from 1960 to 1964 and an associate professor from 1964 to 1969. From 1970 to 1975 he was at the Brooklyn Polytechnic Institute where he served as professor of systems engineering and Director of the Systems Engineering Program. He served as the chairman of the Polytechnic Institute Committee that developed a new B.S. degree program Social and

Technical Sciences using Sloan Foundation funding. This degree program was approved by New York State in 1973. He was instrumental in developing a new multidisciplinary course "System Design in Societal Problems," at the Brooklyn Polytechnic. From September 1975–1977 he served as the director of office of University Research at the U.S. Department of Transportation (DOT) where he was responsible for soliciting, reviewing, contracting, and monitoring over US$4 million dollars of annual research at 60 universities each year. He received the DOT secretary's silver medal in 1977 for his dedicated service. From September 1977–1985 he served as the chairman of the Mechanical Engineering Department at Wayne State University, Detroit, Michigan. He founded the Ford/WSU electronics and computer controls re-education effort involving an M.S. degree program, two separate certification programs, and a seminar/symposium program. From 1985 until he left Wayne State in January of 1987 he was the associate dean for engineering research and graduate programs. From 1987 to 1989 he served as the head of Mechanical Engineering at Texas A&M University. From 1989 to 1991 he was a Texas Engineering Experiment Station (TEES) research professor. From 1991 until his retirement in 1999 he was a professor of mechanical engineering at Texas A&M University, where he developed the engineering ethics program.

Mike served on a number of important committees and often served as chair of many. While at Wayne State University he served as the chair of the transportation committee of the University Cultural Center Association. His tenure on this committee spanned from December 1977 to January 1987 and produced a US$2 million transit system with free service for local transportation with SEMTA, D-DOT, and private support. Service was initiated in January 1990 and is still running. He served as the task force chair for a U.S. Department of Energy study, "Process and Systems Dynamics and Control," resulting in a US$13.5 million program.

In 1981 Mike served as a member of the NSF/Mechanical Systems Steering Committee to identify NSF research priorities. Emphasis in control systems, acoustics, and dynamic systems in civil and mechanical systems was in large part due to the report prepared by the task force on behalf of the ASME Board. Later in 1985 Mike chaired an NSF funded workshop administered by ASME to identify research needs for a new NSF program, "Design Theory and Methodology." The recommendations from the workshop formed the core of NSF's design program in the mid 1980s. In 1983 he served as a member of the DOE/ORNL select panel to identify research priorities in intelligent machines. Later he served as chair of the Oak Ridge National Laboratory/ Center for Engineering Systems Advanced Research Advisory Committee.

I remember Mike as a truly inspiring teacher when I first studied control systems from him in 1979. I vividly remember how he used to get excited talking about feedback and the need for super-accurate sensing within the system bandwidth. My own research inspired by this exposure helped me better understand the subtle ramifications of this statement. In addition to being an outstanding teacher, Mike coauthored two textbooks in control, two textbooks on engineering ethics, and a monograph. Among these are *Control and Dynamic Systems* coauthored with Takahashi and Auslander (Addison Wesley, 1970), *Introducing Systems and Control*, (McGraw Hill, 1974), and *Engineering Ethics: Cases and Commentaries* (Wadsworth, 1995). In the later years Mike fell in love with engineering ethics and took a keen interest in the subject. He wrote about the interplay between control education and ethics in several archival publications.

Mike was an active member of the education panels of ASME and IFAC. He chaired the ASME education panel from 1990 to 1993 and was the chair of the IFAC Education Committee from 1990 to 1993. Mike's dedication to education was so deep that conference sessions on education became a feature at major control conferences due to his efforts. In recognition of Mike's broad and lasting contributions to control education, he was awarded the American Automatic Control Council's John R. Ragazzini award in 1991. He later received the ASME's DSC Award for Education. He was the first recipient of the DSCD biannual leadership award in 1996. It was a fitting tribute to the leadership that Mike provided to the DSC for many decades.

Mike was elected a fellow of ASME in 1974. He received the ASME Centennial Medal in 1981 for substantial contributions to the DSC and the society and a special ASME services award in 1985. Mike held visiting faculty positions at Jadavpur University, Calcutta, India in 1966, University of California, Berkeley in 1966–1967, University of Tokyo, Japan in 1971, and Laboratoire d'Automatique, Polytechnic Institut de Grenoble, France in 1975.

Mike loved classical music, symphony, and the opera. He was an enthusiastic sports fan. He even tried to learn the complex rules of the game of cricket every time he made a visit to the United Kingdom. He loved to eat out and always knew the best restaurants in town. Mike was a family man. He was a devoted husband to his dear wife, Joan. He was an affectionate father to his children, Andy, Evan, and Alix. A memorial service was held in College Station, Texas, on October 5. In recognition of Mike's many lasting contributions to the Dynamic Systems and Control Division of the ASME, the biannual DSC Leadership Award was named after Mike and will now be known as the Michael J. Rabins Leadership Award. Mike will continue to influence our activities in years to come, and he will be sorely missed by his family, friends, mentees, and colleagues all over the world.

Suhada Jayasuriya
Department of
Mechanical Engineering
Texas A&M University

To be included in the conference calendar, send announcements to:

John Watkins
j.watkins@ieee.org

▲ Indicates CSS-sponsored conference
● Indicates CSS-cosponsored conference

The complete and current list of CSS-sponsored and cosponsored conferences is available at the CSS Web site, http://www.ieeecss.org.

» 2008

UKACC CONTROL CONFERENCE (CONTROL 2008)
2–4 September, Manchester, U.K.
General Chair: Hong Wang
Program Chairs: John O. Gray, Guoping Liu
http://www.control2008.org/index.php

▲ **MULTICONFERENCE ON SYSTEMS AND CONTROL**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzales
Program Chairs: Gary Balas (CCA), Marco Lovera (CACSD), Kevin L. Moore (ISIC)
http://conferenze.dei.polimi.it/msc08/

▲ **CONFERENCE ON CONTROL APPLICATIONS**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzalez
Program Chair: Gary Balas
http://conferenze.dei.polimi.it//msc08/cca/cca_index.htm

▲ **COMPUTER-AIDED CONTROL SYSTEMS DESIGN**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzalez
Program Chair: Marco Lovera
http://conferenze.dei.polimi.it//msc08/cacsd/cacsd_index.htm

▲ **INTERNATIONAL SYMPOSIUM ON INTELLIGENT CONTROL**
3–5 September, San Antonio, Texas, USA
General Chair: Oscar R. Gonzalez
Program Chair: Kevin Moore
http://conferenze.dei.polimi.it//msc08/isic/isic_index.ht

● **INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION, AND SYSTEMS 2008 (ICCAS 2008)**
10–14 October, Seoul, Korea
General Chair: Sung Kwun Kim
Program Chair: Doo Yong Lee
http://2008.iccas.org/

2008 DYNAMIC SYSTEMS AND CONTROL CONFERENCE (DSCC 2008)
20–22 October, Ann Arbor, Michigan, USA
General Chair: Galip Ulsoy
Program Chair: Eduardo Misawa
http://www.dsc-conference.org/

▲ **47TH IEEE CONFERENCE ON DECISION AND CONTROL**
9–11 December, Cancun, Mexico
General Chair: Chaouki Abdallah
Program Chair: Thomas Parisini
http://control.disp.uniroma2.it/CDC08/

● **10TH INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION, ROBOTICS AND VISION (ICARV 2008)**
17–20 December, Hanoi, Vietnam
General Chair: Y.C. Soh
Program Chair: C. Wen
http://www.icarcv.org/2008/

» 2009

▲ **2009 AMERICAN CONTROL CONFERENCE**
10–12 June, St. Louis, Missouri, USA
General Chair: K. Hoo

» FOCUS ON EDUCATION *(continued from page 117)*

research—but, it seems, research was unavoidable; it's my nature. May research bring you these same joys.

ACKNOWLEDGMENT
SCRIPTOR HVIVS ARTICVLI CORDE IPSO DIONYSIO ELECTRO MVLTIBVS EMMENDATIONIBVS GRATIAS AGIT.

REFERENCES
[1] W.C. Faulkner, "Nobel speech," in *The Faulkner Reader*. New York: Random House, 1954; reissue: The Modern Library (a division of Random House) 1977.

[2] M.D. Shuster, "In my estimation" *J. Astronautical Sci.*, vol. 54, nos. 3–4, pp. 273–297, July–Dec. 2006.

[3] M.D. Shuster, "The arts and engineering," *IEEE Control Syst. Mag.*, vol. 28, pp. 96–98, Aug. 2008.

[4] W. Strunk, Jr., E.B. White, and R.R. Angell, *The Elements of Style*, 4th ed. New York and London: Longman, 2000.

[5] *The Chicago Manual of Style*, 15th ed. Chicago, IL.: Univ. of Chicago Press, 2003.

AUTHOR INFORMATION
Malcolm Shuster (mdshuster@ comcast.net.) is director of Research for Acme Space Company. After a decade of research in theoretical Nuclear Physics, he wanted to do something different and spent the next 30 years mostly in the aerospace industry, where he quickly found himself doing research again. More extensive author information can be found in a previous article in the *IEEE Control Systems Magazine* (August 2008). He can be contacted at Acme Space Company, 13017 Wisteria Drive, Box 328, Germantown, MD 20874; http// home.comcast.net/~mdshuster.

Send dissertations in the format shown to:

Dr. John Watkins
Department of Electrical
  and Computer Engineering
Wichita State University
1835 Fairmount
Wichita, KS 67260 USA
+1 316 978 6336
Fax: +1 316 978 5408
j.watkins@ieee.org

"ESTIMATION AND CONTROL WITH RELATIVE MEASUREMENTS: ALGORITHMS AND SCALING LAWS"

Prabir Barooah
University of California, Santa Barbara, Electrical and Computer Engineering, USA
Date: September 2007
Supervisor: João P. Hespanha
Current Address: 322 MAE-A, Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611, USA; pbarooah@ufl.edu; http://humdoi.mae.ufl.edu/~prabirbarooah/publications/BarooahThesis.pdf

"ROBUST ADAPTIVE CONTROL FOR UNMANNED AERIAL VEHICLES"

Nazli E. Kahveci
University of Southern California, Los Angeles, Electrical Engineering Department, USA
Date: December 2007
Supervisor: Prof. Petros A. Ioannou
Current Address: Ford Research and Advanced Engineering, Ford Motor Company, 2101 Village Road, MD 2036, Dearborn, MI 48121-2053 USA; nkahveci@ford.com; http://digarc.usc.edu/search/controller/view/usctheses-m958.html

Failed Leadership

Engine development continued apace after Watt. Engineering of components improved, so that parts fitted together better, and friction between components was reduced. Engines were made larger and more powerful; the governor size increased also, to better manage the larger throttle valves of the larger engines. The engine-operating speeds increased, and this allowed (indeed, required) a reduction in the size of engine flywheels.

There was a problem. The steam engines were not responding quickly enough to changes in load. Sometimes the response was very slow indeed. Worse than that, sometimes the oscillations did not die down at all, and even increased in amplitude and became irregular. A disturbance in the engine speed would no longer be regulated smoothly. It was as if a puppy on a leash had grown into a powerful dog too strong and willful for his master, who lost control. The governor was no longer governing. This problem, of the engines not responding correctly to a disturbance from equilibrium, was given a name: hunting. Perhaps this name originated from the idea of a pack of hounds, when thrown off the scent. Instead of functioning together, and moving in the same direction, the hounds would begin to wander far and wide, trying to pick up the scent once more. This image describes perfectly the seemingly random wanderings of the engine speed (and flyball angle) of a mid-nineteenth-century steam engine that had suffered a disturbance.

—M. Denny, *Ingenium: Five Machines That Changed the World*, Johns Hopkins University Press, Baltimore, MD, 2007, pp. 146–147.

Essential Building Blocks

Artwork by Haoyun Fu