The Signal Processing Information Base

DON H. JOHNSON AND PATRICK NABEEL SHAMI

The impact of signal processing research and development projects is, to a great degree, hampered by long publication delays, ineffective communication of results within the signal processing community, and the lack of publicly available data— measurements of real-world signals rather than simulations— for evaluating algorithms. A sequence of facts and events has brought these problems into focus.

- The *publication delay*— the time from paper acceptance to publication— stands today at about one year. The Society's AdCom has placed this problem at the top of its priority list, but the queue is decreasing surprisingly slowly.

- Few know of the Signal Processing newsletter disseminated electronically from Georgia Tech (sidebar 1). A typical newsletter contains previews of journal contents, workshop announcements, and other "news" items for the signal processing community.

- At the most recent Digital Signal Processing Workshop, two researchers "discussed" the relative merits of their algorithms. When asked if a common set of data was available for comparing algorithms, they said no, but wished it was.

Other research communities have had similar problems, and have found at least a partial remedy in making effective use of the Internet and associated computer networks. After discussions with Dr. John Cozzens, Director of the Circuits and Signal Processing Program at the National Science Foundation, he agreed that the community would be served by an easily accessible, publicly available repository of information: data, software, papers, bibliographies, addresses. Toward this end, the Foundation funded, with AdCom's endorsement, a proposal to establish *SPIB*— Signal Processing Information Base— at Rice University. This information base will provide "one-stop shopping" for both the signal processing researcher and the applications engineer. It represents the modern alternative to and extension of the DSP Program book: Data, programs, and papers can be accessed immediately via Internet and the information base will be continually updated (out-of-date material removed as well as new material added). In this way, *current* research results and interesting (challenging) datasets important to the signal processing community can be sustained over the long term.

We intend for *SPIB* to provide one of the cost-effective means of strongly impacting signal processing R&D recommended in the *Keystone Report on Signal Processing*, p. 6:

"The committee also identified several areas where they believed that a moderate investment of energy and money could have a significant immediate positive effect on theoretical and practical signal processing work. These include the establishment of some sort of community-wide, easily acces-

The E-LETTER on Digital Signal Processing} is issued bimonthly (more frequently in the future) and edited by Professors Douglas B. Williams (dbw@eedsp.gatech.edu) and Vijay K. Madisetti (vkm@eedsp.gatech.edu) at Georgia Tech. Items to be included in the e-letter may be sent to either editor. The editors welcome information from as diverse a group as possible. The e-letter's focus is primarily research and academic issues in digital signal processing and telecommunications. The newsletter covers items of interest to the signal processing community.

Articles are welcomed on:

- Conferences: Programs and Calls for Papers
- Funding opportunities
- Thesis abstracts
- Job opportunities (primarily academic)
- People on the move
- Reports on meetings and workshops
- Recent books and publications
- Software releases (non-commercial)
- Any other items that are of interest to the signal processing community

A recent issue's table of contents consisted of

Awards and Honors
IEEE Signal Processing Society 1992 Awards

Journal Table of Contents
IEEE Transactions on Signal Processing: August, September, October 1993
IEEE Transactions on Speech and Acoustics Processing: October 1993
IEEE Transactions on Communications: July 1993
IEEE Transactions on Communications: August 1993
Circuits, Systems, and Signal Processing

Special Issues of Journals
International Journal of Robust and Nonlinear Control: H and Robust Estimation

Conferences and Calls For Papers
1994 International Conference on Acoustics, Speech, and Signal Processing
First IEEE International Conference on Image Processing
Third International Conference on Automation, Robotics, and Computer Vision (ICARCV'94)
EUSIPCO'94: Special Session on the Higher Order Statistics
IEEE Signal Processing Workshop on Higher-Order Statistics

Book Announcements
Speech and Audio Coding for Wired and Network Applications edited by B.S. Atal, V. Cuperman, and A. Gersho

Electronic Technical Report
Toeplitz and Circulant Matrices: A Review by R. M. Gray

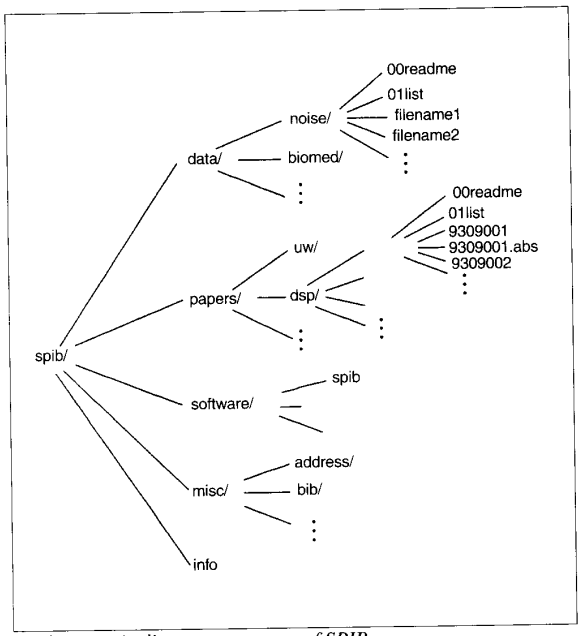
Thesis Abstracts
Image Processing with Multiscale Stochastic Models, by Mark R. Luettgen, Thesis Advisor: Alan S. Willsky

sible library or repository of contributed signal data, and a library of software and algorithms 'known to work,' such as the old IEEE/ASSP DSP tapes. These, it was believed, would improve the effectiveness of many R&D efforts by removing the need to code and test 'building block' algorithms used in the study of more advanced schemes."

This article describes the *SPIB* project, an overview of the initial contents of the information base, and a discussion of the various ways it may be accessed.

SPIB Overview

Because of the wide variety of information available on *SPIB*, perhaps the best way to present its contents is through a map of the system's prototypical directory tree (Fig. 1). This directory structure organizes some 5 Gb of disk (soon to be expanded to 10 Gb) located at Internet address **spib.rice.edu**. The physical computer is a UNIX workstation, and this directory tree literally corresponds to what you will find. At the top level, five directories form *SPIB*'s major subdivisions. Note that *SPIB*'s contents will change; this description merely expresses plans and what is available at this writing.



1. The generic directory structure of SPIB.

data: Datasets are subdivided into categories and are intended to be equivalent to archivable literature, to be used as standard signals for algorithm evaluation, and to be referenced in articles. Preferred datasets are recordings of actual signals; i.e., integer (or floating point) values produced by A/D converters are stored. Currently available data tend to have long durations and exhibit time-varying characteristics.

Examples of such datasets are factory noise, animal vocalizations, biomedical recordings (EEG, for example), and radar array signals recorded in clutter. Suitability of datasets for the information base is judged by members of the Society's technical committees.

papers: Rather than serving as an electronic journal, the *SPIB* repository of technical reports (unreviewed documents, perhaps submitted for publication) is made available to the entire signal processing community less than one week after transmission to *SPIB*. These papers are grouped according to the Society's broad EDICS categories and can be cross-listed. In contrast to data, new papers should become available almost daily. Interested signal processing researchers will receive weekly notification of newly submitted papers' titles and authors.

software: Representing a modern version of the Digital Signal Processing Program book, program sources (Matlab©, C, Fortran) are stored so that newly published algorithms and signal processing environments can become more accessible to the community. Thus, we hope the signal processing community will rely on *SPIB*, as a viable alternative to journals and magazines, for publishing algorithm and design software.

misc: To provide general information, *SPIB* will contain an address list of signal processing researchers, bibliographies, and other items.

info: All of *SPIB*'s information files describing access and submission procedures are kept in this directory.

Storage formats

In virtually all cases, information on *SPIB* is stored in ASCII. For data, in particular, we chose this method rather than binary fixed-or floating-point representations. Related files are kept together in **tar** archives. To save space, these and single files are then passed through UNIX's data compression program **compress**. Macintosh and PC-compatible versions of **compress** and **tar** are available on *SPIB* in the software section.

Data

All data are stored as **compress**'ed ASCII files. The signal values represent their original form (actual values generated by the A/D converter in the sampling process), or as close to this as possible. All samples taken at the same sampling time should appear on the same line, separated by at least one space. In this way, multichannel data as well as single recordings can be expressed by a common format. This format allows Matlab users to read a data file directly, creating a matrix of recordings. Each data file includes a preamble specifying:

- A brief description (title) of the data
- How the signal was acquired: sample rate, presence and

description of any filters or prefilters, description of pre-emphasis or other types of conditioning applied to the signal

- Where the signal was acquired: as complete a description as possible of the environment in which the signal was recorded
- Creator and submitter of data: how to reference the data, whom to contact for further information.

Papers

Each paper is stored in one of two ways: either the source, with *TeX* and its offshoots (particularly (*LaTeX*) as the standard format for text and PostScript the format for included figures, or the formatted output of the entire paper expressed in PostScript. When transmitted to *SPIB*, each paper is assigned a **paper#** according to the year, month, and relative time of submission; for example, the thirteenth paper submitted in August of 1993 is assigned the number 9308013. The text of the paper is stored in a file named **paper#.**

A paper's figures are merged with the source into a single **tar** file. The preamble of a submitted paper contains, in unformatted text, the paper's title, authors, and abstract. This information is stored in a file named **paper#.abs.** The abstracts of a month's submissions are merged into uncompressed ASCII **list** files for browsing.

Software

We welcome algorithmic, design, and simulation software on *SPIB*. The sources for a given software system will be placed in a single **tar** file and **compress**'ed. We are reluctant to impose language standards, and rely on the programmer to produce portable programs.

As with the DSP Program Book, examples should be included so that the person installing the software can determine if installation errors exist. Review mechanisms will be established for programs to determine their utility.

Accessing *SPIB*

Although each information category has its own peculiarities, information retrieval procedures have been designed to be as similar in "look and feel" as possible. To begin and to obtain detailed instructions on obtaining any of *SPIB*'s entries, send electronic mail to **spib@rice.edu**, with the message's subject containing the single word "help." A program, known as the mailserver, responds to your request, sending instructions on obtaining more specific information.

Generally, indexes and searches are available through electronic mail, **ftp**, and **gopher**. Papers and general information can be obtained through the same means, but data and programs more than likely require use of anonymous **ftp** or **gopher** (see table).

Gopher

Gopher represents an exciting, interactive way of retrieving all kinds of information (sidebar 2). Gopher servers, the one at *SPIB* being but one example, run on computers throughout the Internet. Information about them is obtained by your local **gopher** program (the X-windows variant is known as **xgopher**, which searches throughout the network to satisfy your information requests. Through **gopher**, any of *SPIB*'s entries can be retrieved.

Anonymous ftp

Anonymous **ftp** (sidebar 3) can be used to access *SPIB* in a more tedious and almost as capable fashion when compared to **gopher** (keyword searches cannot be performed in **ftp**). You can use **ftp**'s commands to browse the directory structure, retrieving any interesting files. Because virtually all files are stored **compress**'ed, *binary transfer mode must be enabled.*

Using **ftp** within a shell script can also allow you to request

Internet **Gopher** is a distributed document delivery service that allows users seamless access to various types of information residing on multiple hosts. The information source is generally transparent, with data supplied from world-wide locations as easily as from a local server. **Gopher** accomplishes this interface by presenting the user a hierarchical arrangement of documents and by using a client-server communications model. The Internet **Gopher** Server accepts simple queries, and responds by sending the client a document or stream of data. Typical selections available on **gopher** include:

- Current catalogs from libraries across the world
- *Current Contents* abstracts, indexing the latest scientific and engineering publications
- Up-to-the-hour weather forecasts, maps, and satellite pictures
- National Science Foundation award abstracts, publications, and information
- Phone directories of thousands of people (at least) who are somehow associated with **gopher** server sites
- Descriptions and "survival guides" associated with a large number of universities and cities

The **gopher** software for various platforms is available (free of charge) via anonymous **ftp** from **boombox.micro.umn.edu** in the **gopher** directory. **Gopher** is a creation of the University of Minnesota.

Table: Accessing *SPIB*

SPIB Service	e-mail	gopher	ftp
browse directories	✔	✔	✔
search keywords	✔	✔	-
retrieve data	-	✔	✔
submit data	*	-	✔
retrieve papers	✔	✔	✔
submit papers	✔	-	-
retrieve software	-	✔	✔
submit software	*	-	✔

*E-mail must be sent to establish data and software submission procedures.

a paper, **uncompress** and un**tar** the file, format it if necessary, and display it on your workstation screen (Fig. 2). We have written such a program; it is available on *SPIB*.

To connect to *SPIB* via anonymous ftp, use:

prompt% ftp spib.rice.edu
Name: anonymous
Password: username@internet_address

Once connected, you can browse *SPIB*'s directories using ftp's commands (cd, ls, dir, and so on). You can also get any .list files that correspond to directories of interest. To transfer data or papers, be sure to use ftp's binary command.

Electronic mail

Electronic mail (e-mail) is the most pervasive way the technical community accesses the Internet. Thus, we allow access to nearly all *SPIB* services through e-mail. Though it is tedious to send sequences of e-mail messages to browse and acquire information, *SPIB* can be accessed in this way.

Binary files are transferred by the **uuencode** routine. Upon receipt of an e-mail message containing a paper, for example, the user must then **uudecode** the message, then **uncompress** and un**tar** the result. Statistical studies show that the size of the e-mail message is actually about 30 percent of the final file's size.

SPIB Announcements

We have established several information mechanisms to keep the signal processing community informed of what resides at *SPIB* and what new has been added.

- The **list** command available through electronic mail and **gopher** allow users to obtain complete listings of a given directory's contents.
- The **find** command allows for keyword searches within each of the five main categories of information maintained at *SPIB* today.
- Announcements of recently submitted papers will be made roughly every week through the same electronic mailing lists used for Georgia Tech's Signal Processing Newsletter and through the Usenet signal processing bulletin board (**comp.dsp**).

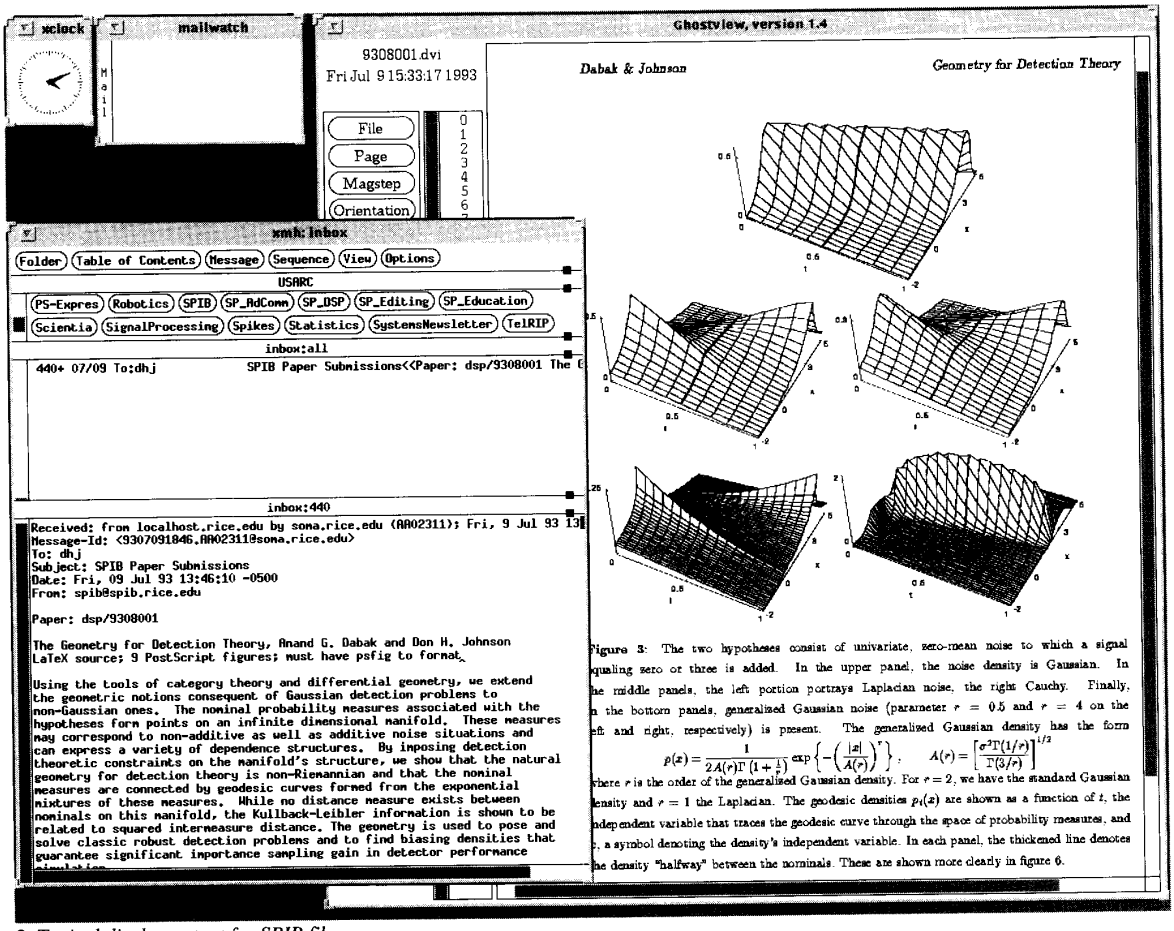
- We intend to establish a *SPIB* column in *Signal Processing Magazine*, detailing updates, presenting access and information statistics, and changes in structure. A portion of this column will be devoted to informing new community members of the fundamental ways to access *SPIB*.

Summary

The information base should provide both applied and research signal processing workers easy access to a vast amount of data, signal processing tools, and up-to-date results. The access tools are standard, allowing anyone with an Internet connection to obtain information. We find **gopher** an exciting new way of obtaining information; in our experience, once you start using it, your workstation becomes an even more valuable tool.

We hope that the data portion of *SPIB* will be especially valuable. Sub-communities within the signal processing field might well settle on particular datasets as particularly challenging. Note the focus on data, not artificial signals. Although testing an algorithm on a small set of signals cannot probe its capabilities, making actual data available for evaluation should make algorithm development more pertinent, even inspiring algorithms that can encompass a wider variety of situations. Furthermore, researchers should be able to mix datasets to create quasi-artificial ones: Noise data can be added to speech in a controllable, easily specified way that other workers can reproduce to test their algorithms. One consequence of this mixing is avoiding the wrangling over defining the signal-to-noise ratio of a given dataset: the resultant SNR could be quoted, but specifying the explicit gains applied to each *SPIB* dataset constitutes a precise description of how the test signal was created.

SPIB will not house all types of data. Because the speech processing community already has a well-established database, only selections of speech data will be maintained at *SPIB* for the general community. Furthermore, images will not be housed at *SPIB*. The image processing community is working to establish an Internet-accessible repository of a variety of images, the size of which would easily swamp *SPIB*'s envisioned 10 Gb storage capacity.However, **gopher** does offer mechanisms for linking databases to appear as portions of one. If **gopher** does become more commonly used, perhaps the various databases should consider unifying under this access mechanism while maintaining separate identities. What we do want are single- and multi-channel data from a wide variety of sources. If you have data of potential interest to the community, don't hesitate to submit



2. Typical display output for SPIB file.

it to *SPIB*. We are establishing a dataset review mechanism through the Society's Technical Committees to reduce redundancy.

Our concept is that data should be as referential as papers, with reference made to the *precise* portion of a particular dataset used in algorithm evaluation. We have written a set of Matlab programs for previewing datasets in elementary ways. For example, short-time Fourier analysis and waveform displays will be provided. In writing these programs, we have adopted a Matlab-inspired convention for referring to data segments: If **filename** is the dataset's name, **filename(m1:m2)** refers to samples $\mathbf{m}_1$ through $\mathbf{m}_2$ inclusively in one-dimensional signals and **filename(m1:m2, n1:n2)** refer to samples $\mathbf{m}_1$ through $\mathbf{m}_2$ inclusively of channels $\mathbf{n}_1$ through $\mathbf{n}_2$.

Although not all kinds of data are maintained at *SPIB*, papers from throughout the signal processing community are welcomed. We have used the Society's EDICS categories to organize papers; this choice does *not* imply that only papers being considered by the Society's various Transactions are welcome. Rather, technical reports and preprints of papers submitted to or accepted by any journalarewelcomesolongastheypertaintosignalprocessing.

We will *not* review papers. In fact, transmitting papers to *SPIB* is limited to electronic mail because software completely handles the files, reading the header, creating the **abs** file, placing each paper in the proper directory, and creating the **list** of recent submissions. We can handle a rather large submission volume this way and should have enough disk space to store papers for the next several years.

One side effect of maintaining a repository for papers in standard formats is imposing a word processing environment (*TeX*) on the community. Informal polls of Associate Editors of the Transactions suggest that a large majority of submitted papers have been formatted that way (the Computer Modern font is quite revealing). Because we do want to serve the entire community, we will support the somewhat less restrictive, but "fatter" PostScript post-formatted output. Pre-formatted source will always be smaller and the same material can be more efficiently stored that way.

However, papers expressed in WYSIWYG formatting systems common on personal computers can also be maintained if they produce PostScript. Over time, we will accumulate statistics and perform surveys to determine which, if any, format the community prefers.

By endorsing the NSF proposal that supports *SPIB*, your Society's elected AdCom representatives have implicitly recognized the community's needs for timeliness and access to data. Although *SPIB* is not an official part of the Society, we will report *SPIB*'s activities to AdCom. If you use the information base, please let us know what you think of it and how it could be improved; send mail to **spib@rice.edu** with a subject line containing the word "comment" and the body expressing your thoughts. To learn how to submit or acquire material, send mail to **spib@rice.edu** with the subject line containing "help."

Acknowledgment

This work was supported by NSF Grant MIP-9301646.

Prof. Don H. Johnson and *Patrick Nabeel Shami* are with the Department of Electrical and Computer Engineering at Rice University, Houston, TX.

Want an engineering or computer science document from any publisher fast?

Announcing **Ask*IEEE**

Why waste time trying to find the sci/tech article you need? Starting January 1, 1993 you can Ask*IEEE ... and get virtually any document in the public domain quickly and inexpensively.

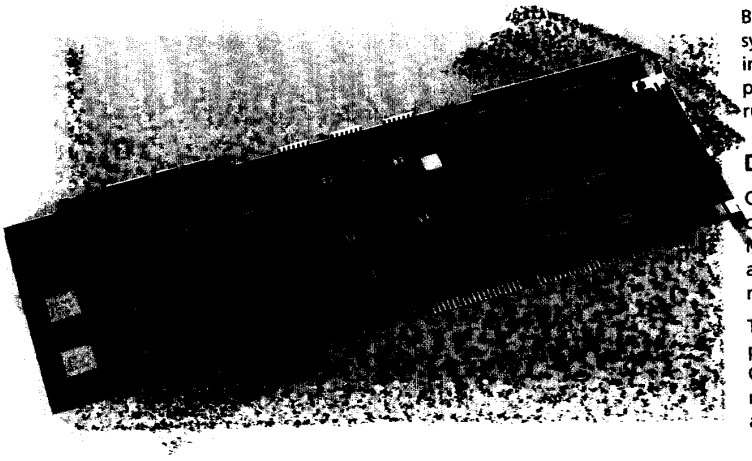
Fast, comprehensive service
The IEEE's immense collection of journals, magazines, conference proceedings and standards will be housed at Ask*IEEE, ready to be delivered to you within hours at no charge.

One-stop shopping at less expense
Rates for articles from the IEEE collection are $12; all other articles are from $15. This simple price will cover all fees, including copyright.

Orders for IEEE items received by Noon PST will be shipped the same day at no extra cost. Delivery will be made the way you choose: fax, e-mail via Ariel fax, overnight express, or mail.

To order, or for more information, contact:
Ask*IEEE
Attn: Manager Document Delivery
P.O. Box 4327 • Burlingame, CA 94011-4327
U.S. & Canada: 1-800-949-IEEE • Rest of world: (415) 259-5040
fax: (415) 259-5045 • DIALMAIL: ASKIEEE • Internet: askieee@ieee.org

DSP Systems for Real-Time Control



The new DS1003 parallel DSP board extends the powerful line of modular dSPACE processor and I/O cards for high-end applications. Based on the TI TMS320C40 DSP, it's 50 Mflops can easily be multiplied by combining more DS1003's or our new four-processor board via the six on-chip byte-wide links. A large number of I/O boards with interfaces for control applications can be connected via a separate high-speed peripheral bus. Programmed in C language, unique software tools for code generation from SIMULINK or MATLAB, for analysis and for DSP monitoring are available now on Sun SPARC and PC/AT.

Development Hardware

Build affordable rapid control prototyping systems with fast DSP and full line of I/O interfaces. With up to 50 Mflops per DSP plus parallelization options you will never run out of speed again.

Development Software

Create your controller with the state-of-the-art design and simulation tools MATLAB™ and SIMULINK™. Interfaces and code generators produce executable real-time code.

Time histories of variables in the DSP program can be analysed in real time. Changing parameters, transferring reference signals and DSP monitoring are easy and efficient.

dSPACE

dSPACE GmbH · An der Schönen Aussicht 2
D-33098 Paderborn · Germany
phone ++49 5251 1638-0 · fax ++49 5251 66529
USA: phone (313) 354-1694 · fax (313) 358-9692

Reader Service Number 26