Numerical and Symbolic Algorithms for Application Specific Signal Processing

Prof. Alan V. Oppenheim

Research Laboratory of Electronics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

Approved for public release; distribution unlimited.

During the period of this grant, our detailed technical accomplishments are reported through journal articles and technical reports. Each of our semiannual reports will highlight certain technical areas and provide a summary listing of our technical articles related to the project.
Semi-Annual Report
Numerical and Symbolic Algorithms for
Application Specific Signal Processing

April 5, 1995–October 30, 1995

Research Organization: Digital Signal Processing Group
Research Laboratory of Electronics
Massachusetts Institute of Technology

Principal Investigator: Alan V. Oppenheim
Distinguished Professor of Electrical Engineering

Grant Number: N00014-93-1-0686
OSP Number: 60314

Program Manager: Mr. Clifford Lau
1 Introduction

During the period of this grant, our detailed technical accomplishments are reported through journal articles and technical reports. Each of our semiannual reports will highlight certain technical areas and provide a summary listing of our technical articles related to the project.

2 Low-Power Signal Processing

We have been investigating various approaches to minimize the power consumption of digital signal processors. Previously we have shown that the best approach to lower power is to reduce the power supply voltage (due to its quadratic influence on the energy) using architectural transformations. Once the supply has been reduced to the lowest possible value (which depends on device threshold voltages, amount of exploitable parallelism in the algorithm, and the available silicon area), the goal is to minimize the switched capacitance.

An important attribute which can be used to minimize switched capacitance is the correlation which can exist between values of a temporal sequence of data, since switching should decrease if the data is slowly changing (highly correlated). For example, time-multiplexing uncorrelated signals on a data bus can result in significant switching activity even though the signals can be independently correlated. Parallel architectures, therefore, generally switch lower capacitance. In many cases, often due limited available silicon area, maximum parallelism is not possible and time-multiplexing is required. The question then becomes, given an algorithm and finite resources, what is the best way to sequence through the operations to minimize switching activity.

To address this question, we first started with a simple FIR filter with length \( L \), assuming that we have one multiplier and one adder available. We assume that the coefficients are stored in ROM and the data in RAM. For an FIR filter, we have to perform \( L \) multiplications inside a given sample period, but the order in which the multiplications have to be performed can be varied. Note that decreasing the number of transitions on the coefficient bus (between the ROM and the multiplier) by re-ordering operations will decrease correlation on the data bus hence increasing transition activity. Though, the optimal ordering to minimize the number of transitions can be determined through exhaustive search for small values of filter length, the
problem grows exponentially for large values of L. Efficient heuristics are therefore required. One such is ordering coefficients from least to greatest. We have verified empirically (for certain values of filter length and word width) that the least to greatest ordering, more often than not, results in fewer expected transitions than an ordering which is chosen at random. The actual percentage change in power consumption is strongly dependent on the values of the filter coefficients.

We are currently evaluating the effect of re-ordering coefficients on the internal switching activity of the multiplier (note that multipliers implemented in static logic can have multiple transitions inside a clock period due to races and the number of extra transitions is a strong function of signal statistics). Based on the results obtained for a single execution unit, we will extend the results to "optimum" sequencing for multiple units (execution and busses). Such an algorithm can be incorporated into tools for assignment and scheduling of dataflow graphs to minimize power.

3 Digital Wireless Communications

One of many important application areas for the approximate processing notions we have been developing is in the area of digital wireless communications. In particular, in limited-bandwidth multiuser systems for communication in environments subject to time-varying multipath fading, maximizing capacity requires a great deal of computationally expensive error-correction coding. This coding is used to combat fading, additive noise, and interference—both co-channel and any jamming. However, in recent work we have shown that a new class of code-division multiple-access (CDMA) systems can dramatically reduce the computational complexity required to achieve a given level of system performance. This is achieved by shifting a greater burden of the job of combatting fading from the error correction coders (and decoders) to the signature sequences used in the system. The key approximate processing notion in this case is that although the complexity required to actually achieve capacity remains the same, the complexity required to come close to capacity in an appropriate sense is significantly reduced. These new systems have a number of significant new features, have important implemenental implications that have only begun to be explored, and may lend themselves to efficient successive refinement receivers. In the remainder of this update, we summarize the key features of these new systems.
In a typical CDMA system for multiuser communication in fading environments, each user's symbol stream is first coded using a suitable error correcting code, then modulated onto a generally distinct signature sequence. With this approach coding does part of the job of combatting the effects of fading and interference among users, while the frequency diversity obtained via the bandwidth expansion inherent in the use of CDMA is also used to combat these effects.

From this perspective, coding is a highly efficient but computationally expensive nonlinear approach to combatting fading and interference, while linear signature-modulation is a computationally inexpensive but somewhat less efficient approach to combatting these effects. As a result, since typically the relevant constraints include not only delay, rate, bandwidth, and power, but also computational complexity, some combination of the two approaches is used in practice.

With traditional CDMA systems, the signature modulation process involves nonoverlapping use of signatures. In particular, each symbol of a user's stream is modulated onto a signature sequence, and the modulated symbols occupy disjoint time slots. However, in this work we describe efficient strategies based on the use of longer signatures that are used in an overlapped manner. We have shown that the additional degrees of freedom can be exploited to dramatically improve the effectiveness of signatures in combatting fading with only a relatively small increase in computational complexity without requiring additional power. Furthermore, unlike conventional CDMA systems, benefit can be achieved without having to incur bandwidth expansion. However, these new CDMA systems, which we refer to as "spread-signature" systems, can also be designed to efficiently exploit any additional bandwidth that may be available.

With these systems, we have shown several asymptotic properties (i.e., for infinitely long signatures) that provide useful insights, and show that good approximations to these characteristics can be achieved in practice. For example, we show that the use of suitably designed orthogonal spread-signatures asymptotically transforms the collection of channels seen by the individual symbol streams from a collection of coupled Rayleigh fading channels into a uncorrelated collection of identical nonfading simple white marginally-Gaussian additive noise channels. In essence, spread-signature CDMA converts various degradations due to fading, co-channel interference, and receiver noise into a single, comparatively more benign form of uncorrelated additive noise that is white and quasi-Gaussian. Furthermore, from the point of view of the channel, the transmitted stream is asymptotically
transformed into a white, marginally-Gaussian process.

In addition, we develop a practical class of binary-valued orthogonal spread-signatures that are particularly well-suited to these applications and have several attractive characteristics. We also show that an optimum class of equalizers for such systems are minimum mean-square error type equalizers, and that with such equalizers substantial performance gains can be achieved over conventional CDMA systems. In particular, given a constraint on the total computational resources available for both coding/decoding and signature modulation/demodulation, the use of spread-signature CDMA can significantly reduce the power required to achieve a target bit error rate while meeting a delay constraint. At the same time, several new engineering challenges are also inherent in the use of spread-signature systems, which we also discuss. A detailed development of these new CDMA systems is presented in [1] and [2], or [3].

4 References


5 Algorithm-Based Fault Tolerance

Our recent work in the area of Algorithm-Based Fault Tolerance has followed up on certain directions exposed in the thesis [1]. (This thesis, developed entirely within the RASSP program, was awarded the MIT EECS Department's top prize for M.Eng. theses, the Guillemot Award.) A high priority has been to make connections between the abstract theory — of fault tolerant computation in various algebraic structures — and actual fault tolerant hardware realizations of algorithms. Using sorting algorithms as an example, we have been able to illustrate how such a connection may be made. Sorting is an important element of various nonlinear signal processing tasks, such as median filtering. Our treatment of the sorting example may serve as a paradigm for other applications, so we briefly summarize it here.

The operation of sorting may be given a semigroup interpretation as follows. Let the members of the "sorting semigroup" $S$ be finite-length integer sequences (possibly including the symbols ",-inf" and/or "+inf", regarded as the minimum and maximum elements). We choose integers for simplicity, but other ordered sets with minimum and maximum elements may be used as easily. The (associative) binary operation on this semigroup takes two such finite sequences and produces a single merged sequence that is sorted from smallest to largest (with all repetitions of elements retained, so the length of the sorted sequence is the sum of the lengths of the individual sequences). In this setting, the sorting of a single sequence of interest could be represented as composition with the null sequence.

To bring fault tolerance into this setting, we need to introduce redundancy by mapping the semigroup $S$ to a larger semigroup $H$ via a semigroup homomorphism $f$. A very simple homomorphism is one that just brackets a sequence $s$ in $S$ by the minimum and maximum elements: $f(s)=-inf,s,+inf$. The semigroup operation on $H$ will be defined as the one that merges two sequences in $H$ into one sorted sequence, but with a single -inf and a single +inf stripped from the result.

What the preceding homomorphism suggests for a fault tolerant hardware implementation is the inclusion of channels for -inf and +inf in the structure. This is easily accomplished in, for instance, an odd-even transposition sorting network, which is a favored parallel sorting approach (on account of the high regularity and low wiring complexity of its VLSI implementation). Some pruning of the design can be carried out to eliminate parts of the hardware that are only trivially exercised by the -inf and +inf. The remaining structure has $N+2$ stages instead of $N$, for sorting a sequence
of N elements. A structure of this type has been described in [2], although we found this reference only after independently arriving at the structure ourselves in the framework of our algebraic theory.

From the above example, one can extract the steps that could constitute a systematic approach to developing ABFT hardware:

1. identify the algebraic structures underlying the computations of interest;

2. explore homomorphisms that introduce redundancy in controlled ways;

3. examine hardware structures that can carry out the redundant computation associated with each type of homomorphism considered in the previous step;

4. trim and adjust the candidate hardware designs, based on criteria specific to each hardware implementation and its associated fault model.

We are optimistic that this paradigm will be fruitful, and work in this direction will continue.

Taking another tack in addressing eventual connections to actual hardware implementations, we have been examining the structure of "machines" that implement group or semigroup computations. In essence, the composition of an input with a present state constitutes the computation, with the next state being the result. There is a rich theory of decomposition for such machines, see [3]. For instance, a group machine can be decomposed into a subgroup machine and a coset leader machine. Correspondingly, a separate code for introducing fault tolerance to a group computation can be seen as introducing redundancy by replicating the coset leader machine.

One benefit of the machine perspective is that it is set up to consider sequences of computations, rather than the individual computations treated in our earlier work. The machine point of view has also yielded valuable insights into some examples of non-separate codes that we have looked at. For instance, it has allowed us to recognize which ways of introducing redundancy are ineffective, in that they only protect against faults that could not have occurred in the original non-redundant structure. We anticipate more general results here.

From the examination of group and semigroup machines in the setting of fault tolerant computation, it is only a small step to considering the notion of fault tolerant dynamics in general. There are many situations in
which the embodiment of a computation is a dynamic system, represented for example in state-space form, or as a Petri net. This is especially true with simulations, or with model-based controllers for dynamic systems. The notion of homomorphic mappings to larger dynamic systems in order to introduce fault tolerance is legitimate here as well, and we have begun to study this. The special case of linear, time invariant state-space models is straightforward to deal with, but the results there are still interesting, largely because the problem does not seem to have been posed this way — the focus of prior work has actually been on the elimination of redundancy (via minimal realizations), rather than on the systematic introduction of redundancy!

We expect to be looking further at protecting computations in semirings. A semiring differs from a ring in that the "addition" operation is a semigroup operation rather than a group operation. An important example is \((\mathbb{Z}, \max, +)\), the set of integers (actually augmented with \(-\infty\)), with \(\max\) being the "additive"operation and \(+\) (i.e., normal integer addition) being the "multiplicative" operation. This particular semiring can actually represent a particular class of Petri nets (so-called marked graphs), and there is a nice matrix calculus that has been developed for it, see [4]. We hope to exercise and extend the semiring results of [1] in the setting of this example.

6 References


7 Publications of Work Supported


ATTACHMENT NUMBER 2

REPORTS AND REPORT DISTRIBUTION

REPORT TYPES

(a) Performance (Technical) Report(s) (Include letter report(s))
   Frequency: Semiannual

(b) Final Technical Report, issued at completion of Grant.

(c) Final Financial Status Report (SF 269)

REPORT DISTRIBUTION

<table>
<thead>
<tr>
<th>ADDRESSEES</th>
<th>REPORT TYPES</th>
<th>NUMBER OF COPIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCIENTIFIC OFFICER CODE: 1114SE</td>
<td>(a) &amp; (b)</td>
<td>3</td>
</tr>
<tr>
<td>Clifford G. Lau</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFFICE OF NAVAL RESEARCH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BALLSTON TOWER ONE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>800 NORTH QUINCY STREET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARLINGTON, VIRGINIA 22217-5660</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADMINISTRATIVE GRANTS OFFICER</td>
<td>(a) &amp; (b)</td>
<td>1</td>
</tr>
<tr>
<td>OFFICE OF NAVAL RESEARCH</td>
<td>&amp; (c)</td>
<td></td>
</tr>
<tr>
<td>RESIDENT REPRESENTATIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ONR MIT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>495 SUMMER STREET ROOM 103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOSTON MA 02210-2109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIRECTOR, NAVAL RESEARCH LABORATORY</td>
<td>(a) &amp; (b)</td>
<td>1</td>
</tr>
<tr>
<td>ATTN: Code 2627</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASHINGTON, DC 20375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFENSE TECHNICAL INFORMATION CENTER</td>
<td>(a) &amp; (b)</td>
<td>2</td>
</tr>
<tr>
<td>BUILDING 5, CAMERON STATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALEXANDRIA, VIRGINIA 22304-6145</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the Scientific Officer directs, the Grantee shall make additional distribution of technical reports in accordance with a supplemental distribution list provided by the Scientific Officer. The supplemental distribution list shall not exceed 250 addresses.