DARPA/Defense Sciences Office
Applied & Computational Mathematics Program
Origins

Louis Auslander (1928–1997)

Prepared for DARPA/DSO by Anna Tsao[1]

[1] This retrospective was made possible because of the contributions of many individuals who gave generously of their time and energy in providing input, feedback, and writing, with special thanks to William Barker, Gregory Beylkin, Dennis Braunreiter, Russel Caflisch, Douglas Cochran, Ronald Coifman, James Crowley, Benjamin Dembart, Jon Ebert, Abbas Emami-Naeini, Fariba Fahroo, George Fann, Franz Franchetti, Kristen Fuller, Leslie Greengard, Mark Gyure, Robert Harrison, Peter Heller, Jeremy Johnson, Robert Kosut, José Moura, Arje Nachman, Stanley Osher, Mark Oxley, Vladimir Rokhlin, Leonid Rudin, Randall Sands, Carey Schwartz, Mark Stalzer, Scott Stewart, Bruce Suter, Haydn Wadley, Steven Wax, and Stuart Wolf.

Table of Contents

Preface

Started originally to jump-start mathematical research that would realize the promise of large-scale parallel supercomputers, DARPA Defense Sciences Office's (DSO) Applied and Computational Mathematics[2] Program[3] (ACMP) pivoted a generation of pure and applied mathematicians toward interdisciplinary work that enabled groundbreaking methods and techniques applicable to a wide range of DoD-relevant science, engineering, and technology. ACMP focused on intractable problems where fresh, non-incremental mathematical perspectives were at hand. Projects in diverse arenas produced new technical approaches, research directions, and multidisciplinary communities—resulting in spectacular advances on remarkably short timescales in applications that included computing technology; data analysis; signal processing, compression, and transmission; computational electromagnetics (CEM); and manufacture of advanced materials.

§1 gives an abbreviated summary of ACMP's early years. §2 contains optional reading material elaborating on some of the accomplishments described in §1. §3 briefly mentions extant DoD challenges that would benefit from radically new mathematical perspectives.

1 An Abbreviated History

In the early 1980s, DARPA was making substantial investments in concerted hardware and software research for novel parallel architectures. Developments were sufficiently promising that the potential for addressing real-world problems at scale was very high. However, paradoxically, no DARPA-type funding existed for the pivotal, high-risk mathematical research many believed would be needed to provide the anticipated revolutionary improvements in computational capability. Dr. Louis Auslander (Distinguished Professor, City University of New York Graduate Center)—a prominent mathematician known in academic, industrial, and government circles for diverse achievements in mathematics and applications (e.g., radar, fast algorithms for Fourier transforms)—persuaded the DARPA Director that it was in the DoD's interest to focus explicitly on mathematics. As a result, ACMP began in 1985.

DARPA's willingness to fund high-risk, high-reward research to meet anticipated long-term DoD needs had routinely produced quantum leaps in capability and technology. Its projects were characterized by their single-minded pragmatic focus on achieving ambitious technical goals, involving researchers from diverse disciplines, as appropriate. DARPA projects had given impetus to entire fields (e.g., materials science, computer science [CS], electrical engineering). Given DARPA's distinctly all-encompassing outlook, what was the added value of a mathematics program? The reality was that in many DoD application domains, the manifold upside potential of mathematical and applications researchers collaboratively taking a fresh theoretical look—from the outset—was not even on the radar. Furthermore, an ever-growing chasm between mathematical theory and application largely prevented timely technology transfer, especially in situations where the required technical effort and technology development would be expensive, lengthy, and manpower intensive.

Ultimately, to overcome these obstacles, ACMP pioneered an entrepreneurial approach to nurturing and expediting state-of-the-art mathematical approaches that was a counterpoint to the predominantly hands-off involvement of academic mathematicians in technology-focused research.[4]

---

[2] "Mathematics" refers to all mathematical sciences, and "mathematician" is shorthand for mathematical scientist.

[3] ACMP was a collection of programs and efforts that explored the application of novel mathematics concepts to computation. It was not a standalone program with the same meaning as current DARPA programs.

[4] While mathematicians in industry dealt successfully with application problems on a routine basis, their job demands often precluded considering

A brief description of the academic mathematical research environment at the time may be helpful in appreciating ACMP's divergence from the status quo. After World War II, the number of research mathematicians in academia grew substantially, but the fundamental research questions considered were increasingly insular and untethered to applications. Game-changing mathematical theories were frequently overlooked because the results obtained were not disseminated outside the self-contained, specialized fields involved in the research; were difficult to even recognize as potentially useful; or had not been developed to the point where they could be readily applied.

Many existing methods, invented when problem scale and computing capability were modest, were unable to meet the stringent scalability,[5] accuracy, and robustness demands of addressing dramatic increases in problem size and complexity. Compounding this situation were the rapid growth and advances in modern fields such as electronics, materials science, and medicine. By the 1980s, many questions of interest to current applications had limited, if any, visibility within the academic mathematical research community. The infrequent communication between mathematicians and researchers in other fields meant that any tech transfer largely occurred—often on glacial timescales—by "throwing mathematics over the fence" or stovepipe approaches.

The prevailing modus operandi and complacency were reinforced by Government-agency practices after Sputnik regarding funding of academic mathematical research. The National Science Foundation (NSF) was (and remains) the primary government funding source[6] for academic mathematics research,[7] with the remaining funding provided by other agencies. Funded research focused almost exclusively on self-directed, curiosity-driven basic research involving single investigators, and merit was primarily judged based on publications showcasing the novelty and positive attributes of theoretical results.

At mission-based agencies, program managers (PMs) funded areas according to relevance to future application needs. Initiative and advocacy by individual PMs and investigators were the main conduit for tech transfer of new theories to relevant application stakeholder communities, and unfamiliar mathematics was often met with resistance or skepticism. Mathematical work related to bottom-line-oriented, real-world applications was largely funded separately within the Government (e.g., engineering, medicine, DoD research at the 6.2 level or greater)—often undertaken long after the original ideas had been published and seeking to avoid risk by pursuing familiar, often-outdated theories. The result was that theoretical deficiencies or gaps frequently were addressed via (sometimes Herculean) one-off approaches.

Auslander was ACMP's second PM during 1989–91.[8] To seek out and nurture opportunities for maximal DoD benefit, Auslander originated speculative research projects[9] in strategically selected mathematical areas. It should be pointed out here out that the amount of ground that Auslander covered in two years was epic by any standard, but particularly so since almost all the projects he initiated were outside his own research areas at the time. However, his approach was in keeping

more fundamental, long-term questions or, in many cases, doing any mathematical research. The fact that mathematicians had success in diverse industrial settings has often been attributed to their approach to problem solving, which transcends subject matter and is somewhat particular to their training.

[5] Algorithms are "scalable" if computational gains grow proportionally with increases in computational resources. Otherwise, increases in computational resources lead to diminishing returns.

[6] R. Malek-Madani and K. Saxe, *Federal funding for mathematics research*, Notices of the AMS, 66 (4), 2019, pp. 576–580.

[7] From 1988–1992, over 75% of Ph.D. graduates from mathematics (excluding statistics or operations research) departments were employed in academia, as reported in *The SIAM Report on Mathematics in Industry*, Jan. 30, 1998.

[8] Some accounts of Professor Auslander's DARPA involvement are given in S. S. Chern; T. Kailath; B. Kostant; C. C. Moore, Coordinator; and A. Tsao, *Louis Auslander (1928–1997)*, Notices of the AMS, 45 (3), 1998, pp. 390–395.

[9] These were in the spirit and on the scale of what are now known at DARPA as "seedlings."

with his belief that when mathematicians actively engage in application areas new to them, both the applications and mathematics itself would be rejuvenated. He engaged in a proactive outward-looking campaign of multidisciplinary recruitment and outreach, involving individuals from the DoD, academic, and applications communities. The resulting multidisciplinary interactions framed application needs and facilitated new collaborations between mathematicians, scientists, technologists, and engineers, some of which resulted in funded projects.

ACMP's overarching technical themes (see §1.1) were germinated by some of these early projects. A key feature of these projects was that rather than just trying to restrict attention to a preconceived set of narrow questions, they sought to discover and address the generalizable mathematical underpinnings for core questions arising in a broad spectrum of applications. The far-reaching scope of these early trailblazing projects led to numerous, sometimes unexpected, opportunities for high technological impact. Building on the fruits of Auslander's groundwork, subsequent ACMP PMs became forward-looking champions of a concerted, integrated theory-to-practice style of ambitious, speculative, multidisciplinary research initiative. This document describes some notable projects and "programs"[10] that are testaments to the sweeping impact of this pivotal period.

The progression of successful ACMP initiatives from fundamental explorations to subsequent application-oriented programs often required longer periods of investment than was typical for DARPA. DARPA empowered ACMP PMs to undertake mathematical risks and, upon success, to make the significant investments to stay the course in achieving technological impact. ACMP's mode of funding enabled dramatic compressions in the timelines for realizing the fruits of new mathematical developments compared to the usual laid-back approaches.

Projects variously encompassed mathematics, CS, physical science, engineering, and application—with an insistence on involving all stakeholders in framing and addressing the actual technical goals. ACMP PMs employed outreach, educational, and teambuilding strategies adapted to the circumstances, often in collaboration with other DARPA PMs or funding agencies, to forge any needed connections among disparate research communities and disciplines. Typical groundwork preceding pitching and initiating ACMP projects included recruitment of mathematical and applications researchers with relevant expertise and facilitation of cross-disciplinary technical discussions to bridge disciplinary language barriers. This was often an uphill battle since the mathematicians being recruited were often unfamiliar with (or even resistant to) DARPA, had minimal or no previous applied research or computing experience, were unfamiliar with the specific applications of interest, or had to be persuaded that the problems were of mathematical interest to them. Equally challenging was convincing application stakeholders of the potential benefits of exploring unfamiliar mathematical approaches and collaboration with mathematicians. As well, numerous intradisciplinary barriers needed to be overcome (e.g., between pure and applied mathematicians, between theorists and experimentalists, between physical scientists and engineers, between academic and industrial researchers).

As demonstrated over 30+ years of DSO funding, ACMP's enterprising, opportunistic approach produced fundamental mathematical advances that enabled radical capabilities and significant cost savings compared to the then state of the art. Moreover, ACMP collaborations radically changed how problems were defined and approached across many of the disciplines involved, which in turn instigated new research avenues. A notable development was that numerous research mathematicians became involved in small companies pursuing research directions and

[10] "Programs" comprised multiple projects with a common set of broadly defined technical objectives. It only became ACMP practice beginning in the mid-1990s to give programs memorable thematic names; in this document, unnamed programs are identified by their approximate time frame.

tool and technology development initiated by ACMP projects. Another ACMP legacy is the extent to which mathematical considerations and former ACMP performers have become part of the broader DARPA landscape as evidenced by several instances of PM, thematic, and performer crossover, a few of which are mentioned or described in subsequent sections.

From the perspective of mathematics as a discipline, ACMP projects greatly expanded the reach, scope, and boundaries of mathematical theories and approaches; enlarged the pool of academic mathematicians exposed to DoD-relevant problems; fostered collaborative, applied mathematics research communities; and improved the dynamics of mathematical innovation and technology transfer.

1.1 Overarching Technical Themes and (Selected) Technical Accomplishments

Mathematics' power in facilitating understanding of real-world phenomena begins with "effective" representations, that is, abstractions that embody relevant structure and knowledge in a way that accommodates mathematical machinery for rigorous reasoning. Moreover, while there may be many possible representations, "effectiveness" is often highly dependent on context. A tacit assumption underlying mathematical research is that the subject of study has mathematically exploitable "structure." ACMP sought to identify areas where expediting novel fundamental mathematical developments might plausibly result in radical advances in capability; the overarching technical drivers were the aforementioned effective mathematical representations and fast, scalable algorithms.

ACMP projects sought to exploit domain-specific structure and behaviors inherent in physical systems and phenomena. In ACMP's formative years, projects evolved naturally into four discernible discipline-independent themes: (1) data distillation; (2) analysis-based fast algorithms; (3) prediction, design, and control of physical processes and systems; and (4) architecture-aware[11] algorithmic representations. An overview of each theme is presented in §1.1.1–§1.1.4 and illustrated with a few notable exemplars.[12] For interested readers, §2 contains additional technical details. Since this document only details mathematical directions initiated prior to this century, each subsection concludes with a list of some later-occurring thematic ACMP reference-point projects/programs.

Given the abundance of potential applications, ACMP PMs could turn their attention to the most promising avenues. Over time, considerable cross-fertilization occurred among ACMP themes and with other application domains, DSO PMs, DARPA offices, and other funding agencies.

1.1.1 Theme 1: Data distillation

The need to acquire, analyze, and exploit data is omnipresent, and for many DoD-relevant digital technologies (e.g., signal/image processing, communications), processing demands have always exceeded existing computing capability. Furthermore, such shortfalls have been exacerbated by suboptimal algorithms and small-footprint requirements. This theme focused on novel mathematical methods to eliminate impediments to timely, accurate processing of data sets. Early projects mainly addressed signal processing applications, new modeling approaches, and novel approximation methods. Many of the approaches have become computational mainstays of modern communications, data/signal/image processing and analysis, and numerical algorithms and are being adopted by machine learning (ML) and artificial intelligence (AI) practitioners.

[11] as this type of approach came to be known later

[12] Most of the technical content herein was derived from project vignettes written by former ACMP performers based on exchanges between the author and the performers.

One pioneering methodology resulted from projects that developed an entirely new mathematical methodology for denoising digital images by considering a mathematical quantity known as total variation (TV).[13] The researchers made the surprising proposal to adapt TV methods, used heretofore in studying partial differential equations (PDEs), to image processing. In this case, excessive or spurious image detail translates to high TV, and the aim was to reduce it. The resultant PDE model, in contrast to methods being used at the time, preserved edges while removing noise. The resulting breakthroughs led to widespread academic and commercial use of TV-based methods in diverse signal, image, and data processing contexts. As a matter of interest, the recent, first image of a black hole resulted from using this model.

Many projects successfully exploited the multiscale nature of physical phenomena by expanding upon several previously untapped and closely related mathematical disciplines. Over the next few decades, the impetus from these projects instigated revolutionary advances for problems involving large-scale data (e.g., data processing and analysis, signal/image processing, waveform and filter design, navigation, digital photography), signal-processing system design, and large-scale simulation (Theme 2).

By the 1980s, a small number of transforms, led by fast Fourier transforms (FFTs), were the algorithmic workhorses of digital signal processing, but there were no effective general strategies for representing signals having discontinuities or transients or for finite, nonperiodic, or nonstationary signals. Starting about 1989, ACMP projects explored the implications of recent wavelet-inspired mathematical developments and facilitated what many consider one of the fastest, most successful examples of mathematical technology transfer.

In the 1970s, a family of wavelets[14] was created to address some of the same types of problematic signals in geophysics, followed by a burst of research in "wavelet analysis" in the geophysics community. After the observation in the 1980s of the strong parallels between wavelet analysis and a substantial body of prior work in pure math, a period of significant mathematical developments ensued. Three notable research directions emerged: (1) multiresolution analysis, (2) its digital signal processing counterpart—which was amenable to fast computation, and (3) a budding revolution in subband coding design capability.

However, by 1989, the greater DoD community was still largely unaware of wavelet-inspired developments. Recognizing an opportunity to vector research in practical directions of benefit to DoD, Lou Auslander launched some wavelet-related exploratory projects involving mathematicians and defense contractors. Their success laid the foundation for subsequent, larger-scale programs initiated by multiple PMs aimed at expanded mathematical development and application in a broad variety of DoD-relevant contexts. An astonishing number of academic mathematicians with no previous applications experience played a pivotal role in the ensuing developments, some of which are described here.

ACMP projects (see §2.1) also drove the creation of a world-wide multidisciplinary research and development (R&D) community that covered a vast amount of application ground (e.g., automatic target recognition [ATR], denoising, filtering, signal/image processing, communications) and developed new theories and fast multiscale and time-frequency methods for adaptive waveform design, filtering, functional approximation, and signal resolution that were not amenable to Fourier analysis. Transformative technologies emerged, such as Digital Subscriber Line (DSL) and wavelet compression of digital imagery—perhaps the best-known achievement of ACMP funding.

[13] The original TV paper, L Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60, 1992, pp. 259–268, has 16,350 citations on Google Scholar.

[14] The first wavelet was invented by Alfred Haar in 1909.

Wavelet compression has been adopted in official standards for compression (e.g., fingerprints, medical images, and general still images and motion capture) and is widely used today to preserve superior image quality at high (lossy) compression ratios (small file sizes), while offering smooth transitions at intermediate file sizes.

The impact of numerous methods developed because of ACMP funding has only grown over time, as the theory and tools continue to evolve and influence the state of the art in an increasing number of fields including computational science, physical science, medicine, data science, and engineering. However, perhaps ACMP's most far-reaching technical legacy in this area is the heightened awareness of the ubiquity of the need for application-appropriate multiscale mathematical representations and tools. DSO's ACMP was uniquely placed to support these technical developments, because they occurred as a synthesis of ideas from pure mathematics, signal processing, and image science. By funding practitioners in these disciplines to communicate and innovate together, a body of knowledge and robust technical standards were developed that could not have come from any one field alone or from within the confines of a product-focused commercial enterprise.

Subsequent thematic programs include ACMP's Integrated Sensing and Processing (ISP), Topological Data Analysis (TDA), Waveform Adaptive Sensing, and Geospatial Representation and Analysis (GEOSTAR) as well as DARPA's Microsystems Technology Office (MTO) Analog-to-Information (A-to-I) and Multiple Optical Non-Redundant Aperture Generalized Sensors (MONTAGE) programs, to name a few. In some of these programs, the confluence with Theme 4's architecture-aware perspective led to a co-design approach to optimizing overall integrated hardware/software solutions.

1.1.2 Theme 2: Analysis-based fast algorithms

In the 1980s, a dearth of accurate, scalable algorithms was a significant barrier to effective use of large-scale computational resources for simulating physical phenomena in applications such as electromagnetics, quantum chemistry (QC), and gravity. Indeed, existing simulation approaches then in use by the DoD CEM community were unable to provide the needed speed or accuracy to consider realistic-size problems, let alone in realistic contexts.

The dominant issue was the absence of fast, accurate algorithms that would make virtual experimentation possible for DoD applications. For a single case involving a realistic-sized aircraft model, the codes available in the 1980s either ran for days/weeks or ground to a halt, whereas the desired turnaround was in hours per run. Even assuming the existing methods had been sufficiently accurate, relying only on hardware speed improvements due to Moore's law would have required some 10 hardware generations to scale to the desired problem size and turnaround times.

Groundbreaking mathematical work in the 1980s resulted in the scalable Fast Multipole Method (FMM) for performing many-body simulations.[15] Intuitively speaking, FMM makes use of the fact that interactions between groups of particles can be treated in a multiscale fashion (i.e., mutual interactions can be treated with increasing amounts of aggregation as the distance between them grows—with no loss of accuracy) to produce a scalable, parallelizable algorithm using sparse matrices. This revolutionary algorithm paved the way for a new algorithmic paradigm that uses appropriately constructed analysis-based representations of multiscale physical phenomena to devise fast divide-and-conquer algorithms. FMM's underlying mathematical insight showed promise for supplanting a broad class of algorithms then in use.

[15] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comput. Phys., 73 (2), 1987, pp. 325–348; one of the Top 10 Algorithms of 20th Century, IEEE magazine, Comput. Science & Engineering.

After an analogous method in two dimensions was developed for one of the equations that govern electromagnetic scattering,[16] the ACMP PM, Lou Auslander, recognized that a three-dimensional analog[17] could be the first step in eliminating the deficiencies of existing CEM codes intended for use in the design of stealthy military aircraft. Auslander's hope was to divert the CEM code development community from existing dead-end approaches. Given the numerous concurrent challenges, no single academic or industrial group was equipped to create a coherent roadmap for success.

While the academic community was quick to appreciate the breakthrough nature of the early FMM research, it was inaccessible to most in the applied DoD community. Even to mathematically sophisticated CEM code developers, the path from the mathematics papers to practice appeared too uncertain to expend significant resources. ACMP's initial projects, involving mathematicians, physicists, CEM code experts, and computer scientists, first developed the 3D FMM theory in both the low- and high-frequency regimes and demonstrated its use in an industrial aerospace CEM code on model problems. The paper *The Fast Multipole Method for the Wave Equation: A Pedestrian Prescription*[18]—which could only have resulted from such a multidisciplinary collaboration—translated the mathematical formulation of the 3D Helmholtz FMM into an algorithmic version that was understandable and convincing to practitioners.

However, the software was neither widely available nor industrial strength. Moreover, many mathematical issues remained to be addressed to make FMM into a practical tool in the CEM arsenal (e.g., representation of real-world material properties and geometries), and the community was still largely devoted to competing technologies. An aggressive expansion of the original ACMP-funded collaborations, the Virtual Electromagnetic Testrange (VET) program, was undertaken and involved major defense contractors and DoD labs to facilitate tech transfer. As a result, the electromagnetic scattering for an aircraft model can now be computed in hours as opposed to days as in the 1980s. Ultimately, the algorithms and associated new paradigms for representing complex geometries were adopted by the DoD,[19] and FMM-based CEM simulation tools are now relied on heavily within the DoD community and continue to be developed for new application regimes. More technical details are given in §2.2.

A group of mathematicians working on the ACMP Virtual Integrated Prototyping (VIP) program (Theme 3) began building the theoretical foundation for novel analysis-based approaches to QC computations important in materials science applications. Research continued well after the project, as did the collaboration between these mathematicians and Oak Ridge National Laboratory (ORNL) scientists (with expertise in computational math, chemistry, and high-performance computing). A decade later, this team spearheaded in a mini revolution in quantum simulations. The resulting software, MADNESS (Multiresolution ADaptive Numerical Environment for Scientific Simulation), which employs high-performance parallel algorithms and implementations with

[16] V. Rokhlin, *Rapid solution of integral equations of scattering theory in two dimensions*, J. Comput. Phys. 86, 1990, pp. 414–439.

[17] 3D is generally significantly more challenging than 2D.

[18] R. Coifman, V. Rokhlin, and S. Wandzura, *The fast multipole method for the wave equation: a pedestrian prescription*, IEEE Antennas and Propagation Magazine, 35 (3), 1993, pp. 7–12.

[19] through the DoD High Performance Computing Modernization Program

controllable precision, is now considered one of the most accurate in this field.[20, 21]

MADNESS was recognized by the R&D 100 Awards in 2011. It is an important code to Department of Energy (DOE) supercomputing sites and is used at DOE leadership computing facilities to evaluate the stability and performance of their latest supercomputers. It has a world-wide user base in Government, universities, and industry. More technical details are given in §2.3.

The impact on QC and beyond began in the mid-to-late 1990s with ACMP projects that encouraged multidisciplinary, potentially highly synergistic collaborations. At that time, the high-risk conjecture that analysis-based fast algorithms for QC were possible was based on mathematical intuition. The eventual success was made possible through the knowledge imparted by materials scientists and computational chemists of the underlying application needs and the state of the art.

The generality of analysis-based approaches for efficiently representing multiscale physical phenomena demonstrated in diverse ACMP projects raised awareness worldwide. Consequently, the community working in related algorithmic research evolved from a handful of U.S. researchers to widespread international use and development (e.g., in acoustics, astrophysics, fluid dynamics, electromagnetics, and elasticity). The approach has revolutionized a broad swath of computational science, and instances can be found in commercial electromagnetics and acoustics software (Ansys, Coustyx, European Aeronautic Defence and Space Company [EADS], FEKO [Altair Engineering], Integrand Software/Cadence) and commercial QC software (Gaussian, Schrodinger).

Two other notable thematic projects initiated in the 1990s concerned automated optimal filter design (see §2.4) and gravitational field calculations (see §2.5). In both cases, the innovative, but nascent, theories have only recently begun to have impact within the DoD community after some years of additional research and Small Business Innovation Research (SBIR)/Small Business Technology Transfer (STTR) program funding from outside DARPA. The automated optimal filter design research has produced capabilities well beyond the initial ACMP-funded research, and some of the algorithms have been incorporated into the X-Midas portable, networked, interactive environment for signal processing and analysis. In the case of the gravitational field calculations, the United States Space Force (USSF) adopted a software code[22] based on a model originating in an ACMP project that resulted in a 60% speed up over legacy methods, as well as higher accuracy.

1.1.3 Theme 3: Prediction, design, and control of physical processes and systems

Modeling and simulation were already widely used in many areas of physical science and engineering, and control theory was well-established in many commercial industries. Heavy funding in PDE research (notably, computational fluid dynamics [CFD]) and control theory over the years had resulted in a rich repository of mathematical approaches that had not yet drawn any attention in many applications of DoD import.

---

[20] R. J. Harrison, G. Beylkin, F. A. Bischoff, J. A. Calvin, G. I. Fann, J. Fosso-Tande, D. Galindo, J. R. Hammond, R. Hartman-Baker, J. C. Hill, J. Jia, J. S. Kottmann, M-J. Y. Ou, J. Pei, L. E. Ratcliff, M. G. Reuter, A. C. Richie-Halford, N. A. Romero, H. Sekino, W. A. Shelton, B. E. Sundahl, W. S. Thornton, E. F. Valeev, Á. Vázquez-Mayagoitia, N. Vence, T. Yanai, and Y. Yokoi, *MADNESS: A multiresolution, adaptive numerical environment for scientific simulation*, SIAM J. Sci. Comput., 38 (5), 2016, S123–S142, http://dx.doi.org/10.1137/15M1026171.

[21] S. R. Jensen, S. Saha, J. A. Flores-Livas, W. Huhn, V. Blum, S. Goedecker, L. Frediani, *The elephant in the room of density functional theory calculations*, J. Phys. Chem. Letters, 8 (7), pp. 1449–1457, https://doi.org/10.1021/acs.jpclett.7b00255.

[22] a product of a AFOSR sponsored STTR Phase II project entitled *Innovative Earth Gravity Reformulation and Numerical Integration for Responsive SSA*

One mathematical innovation resulting in part from a 1980s ACMP project was level set methods (LSM),[23] developed in response to numerous applications needing effective mathematical tools for computing interfacial motion (e.g., multiphase, multimaterial problems). This class of methods uses the level set model as the conceptual basis for computational numerical analysis of surfaces and shapes. LSM are an effective tool for capturing geometric properties of the interface, following shapes that change topology, and modeling time-varying objects (e.g., an airbag). Since its invention, LSM have been widely researched and used (e.g., CFD, combustion, trajectory planning, optimization, image processing, biophysics) and played a central role in ACMP's VIP program.

The original LSM paper addressed the 3D motion of an interface that moved normal to itself, at a velocity that was dependent on the surface curvature, which arose in a host of thermo-mechanical-chemical problems of interest to the DoD, such as combustion and materials science. In the mid-1990s, LSM were adapted to the problem of detonation shock dynamics in 3D condensed explosive configurations.[24] By the late 1990s, LSM were being used to represent materials interfaces between internal material parts of 3D components that were being simulated simultaneously in a 3D, dynamic, multi-material simulation. Examples included the propellant combustion interface of the rocket solid propellant, interfaces between metals, and plastics that were subject to rapid and extreme state loading.

LSM, by themselves, are not new at this point, but making them useful for 3D applications with complex interfaces and compatible with the highest resolution possible for the entire simulation domain and its domain subcomponents has been of great ongoing interest since their invention. Two methods of note are the Ghost Fluid Method[25] (GFM-LSM) and LSM applied to the solid rocket motor (SRM) burnback.[26] GFM-LSM has the advantage of being readily parallelized for large-scale 3D computations. SRM-LSM has become the predominant method for representing SRM design and is essential to advanced-concept missile design.

By the late 1980s, complex dynamics, irreversibility, and disparate physical and temporal length scales of materials processes had become dominant obstacles to designing high-yield, cost-effective processing strategies. While there was a robust mathematics community researching the physics of materials growth, there was little research pertaining to materials processing.

Given DSO's prominent role in advancing materials science and processing research, it was only natural that ACMP should explore materials applications. Lou Auslander, who had the vision to introduce modeling, simulation, control, and signal processing methodology to materials processing, persuaded an eminent electrical engineer who "only wrote papers"[27] to participate in a speculative materials processing project involving a real-world system. Its success[28] provided the impetus for ACMP's subsequent heavy focus on electronics-related materials processing. Two

---

[23] The original LSM paper, S. Osher and J. A. Sethian, *Algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys, 79, 1988, pp. 12–49, has around 18,000 citations on Google Scholar.

[24] T. Aslam, J. B. Bdzukm D. S. Scott, *Level set methods applied to modeling detonation shock dynamics*, J. Comput. Phys., 126 (2), 1996, pp. 390–409.

[25] R. P. Fedkiw, T. Aslam, B. Merriman, and S. Osher, *A Non-oscillatory Eulerian approach to interfaces in multimaterial flows (the Ghost Fluid Method)*, J. Comput. Phys., 152 (2), 1998, pp. 457–492.

[26] M. A. Wilcox, M. Q. Brewster, K. C. Tang, D. S. Stewart, and I. Kuznetsov, *Solid rocket motor internal ballistics simulation using three-dimensional grain burnback*, Journal of Propulsion and Power, 2007, 23 (2), pp. 575–584.

[27] Professor Tom Kailath (Stanford); an anecdote about this project appears in *Louis Auslander (1928–1997)*, Notices Amer. Math. Soc., 45 (3), 1998, pp. 390–395.

[28] A paper resulting from that project was the 1994 Outstanding Paper of the IEEE Transactions on Semiconductor Manufacturing, a journal the academic researchers had hardly been aware of at the start of the project.

notable 1990s DSO efforts, Rapid Thermal Processing (RTP) and VIP, were instrumental in introducing model-based control and the successful design, scale up, and control of materials processes involved in the fabrication of atomic-scale transistors and electronic materials.

As transistors shrank, RTP was used to shorten thermal cycles in wafer-processing furnaces to reduce defects. The then state-of-the-art, simple black-box control schemes were simply inadequate because of the complexity of the underlying nonlinear, radiation-dominated physics and the challenges of maintaining dynamic temperature uniformity across wafers of increasingly greater diameters. In the mid-1990s, the use of physics-based models in real-time controllers had never been contemplated, because it was believed that only mathematical models involving large systems of differential equations could be effective, and yet, these would be far too slow for real-time use. The key innovation of the ACMP RTP project (See §2.6) was a methodology for dimensionality reduction of high-fidelity physical models to produce models that were sufficiently accurate and computationally fast for real-time online control.[29] For the first time, equipment makers could simulate closed-loop RTP equipment performance to efficiently make and evaluate design changes.

Because of the ability to develop and implement a physical-model-based real-time controller in software, RTP became the enabling step in the thermal process for creating shallow junction transistors and for growing thin gate oxides in semiconductor wafer manufacturing. Currently, most of the feedback controllers running in semiconductor fabrication plants (fabs) worldwide are based on this work, and RTP is a critical piece of the global semiconductor wafer-manufacturing business.

ACMP was instrumental in taking theoretical insights to commercially viable products. Given the conservatism of commercial companies, this tech transfer only occurred because the funding enabled not only fundamental research but also the development of proof-of-principle software that could be tested in an actual fab.

Since the DoD is heavily reliant on advances in processing capability resulting from improvements in semiconductor technology, the impact of this work has been widespread. Over the past 20 years, this same mathematical methodology has been expanded to many other processes, including model-based control of chemical mechanical planarization and metalorganic chemical vapor deposition (e.g., light-emitting diode [LED] production), which are heavily used in the processing and manufacture of advanced materials. Moreover, the mathematical methodologies developed are broadly applicable in many other application domains.

The VIP Program was truly revolutionary because it was the first large Government program to create multidisciplinary teams of physicists, engineers, and applied *and* pure mathematicians from both academia *and* industry to tackle problems of great importance for electronics, a key DoD industry. VIP's ambitious goal was to combine atomistic- to macro-scale theory, modeling, and simulation across multiple spatial and temporal scales with physical design and prototype development of new fabrication tools. Many of the scientists, particularly the pure mathematicians, were new to DARPA projects, and their efforts led to great innovations resulting from new and enduring collaborations between mathematicians, materials physicists, and engineers. The VIP project was, as well, notably forward thinking in that it had a hardware component to test the computational models.

In the 1990s, communication technologies required increasingly faster devices to process high-

[29] A. Emami-Naeini, J. L. Ebert, D. de Roover, R. L. Kosut, M. Dettori, L. Porter, and S. Ghosal, *Modeling and control of distributed thermal systems*, IEEE Trans. Control Systems Technology, 11 (5), 2003, pp. 668–683.

bandwidth data streams. A variety of semiconductor quantum devices, relying on thin layers and abrupt interfaces, were intensively studied for operation in digital logic circuits operating at terahertz frequencies. Paramount in the successful integration of these devices into high-speed circuits was the ability to control the discrete device characteristics (current vs. voltage behavior) to within very narrow tolerances. Achievement of this goal required precise control of layer thicknesses, composition, and interface roughness.

One VIP team—focused on modeling, simulation, and experimental growth of III-V materials for high-speed electronics by molecular beam epitaxy (MBE)—developed new simulation methods for layer-by-layer growth that bridged the atomistic to continuum length scales and allowed for physically realistic, fast simulation (see §2.7). The new *island dynamics model* developed was multiscale, and its original application was for layer-by-layer growth simulated with novel LSM. LSM were virtually unknown to the materials science community before the VIP Program and have become standard for simulating epitaxial growth.

In subsequent work, this methodology was developed further to model multilayer growth (the formation of mounds). It has been combined with elastic models to model the formation of so-called quantum dots. The new methodology was also the starting point for the development of new adaptive multigrid methods.

The interaction of mathematicians and materials physicists in the VIP effort had a broad impact including a new grid-based numerical method for computation of the multi-electron Schrodinger equation. The code based on this method has been used extensively for determining the properties of gated semiconductor quantum dots and is the best software tool of its kind for simulating semiconductor-based qubits. Many groups throughout the world have collaborated with the developers for quantum dot simulations, and plans are proceeding to make the tool more widely available to the semiconductor qubit research community.

Another VIP team developed integrated atomistic-to-macroscopic computer models suitable for controlling thin-film deposition for giant magnetoresistance (GMR)-derived materials and ultimately facilitated a solid-state revolution in data storage. This modeling and simulation capability successfully traversed the long path from materials physics theory to the actual controller implementation that enabled acceptable uniformity when using radio frequency (RF) diode sputtering. The resultant performance improvements and cost savings made RF diode sputtering the method of choice for large-scale manufacturing and enabled a quantum leap in GMR use in commercial products needed by DoD.

This VIP methodology was subsequently used to devise a different approach to thin film multilayer assembly—biased target ion beam deposition (BTIBD), which was commercialized to produce ultra-smooth metal and oxide films for both magnetic and superconducting tunnel junctions. This class of electronic materials, pioneered by DARPA and now referred to as *spintronics*, resulted in a solid-state revolution in data storage and promises massive increases in computational power. The commercial company involved[30] is now a leader in the practical commercialization of spintronics and manufactures high-performance spintronic products including sensors and couplers that are used to acquire and transmit data. More details are given in §2.8.

VIP is an excellent case in point for illustrating some of the challenges frequently faced by ACMP PMs beyond the usual extensive technical groundwork done by DARPA PMs. Thin film growth was not a well-known or active research area within mathematical circles. Furthermore, the PM's assessment was that to achieve maximal results, collaborations among theoretical, applied,

[30] NVE Corporation, https://www.nve.com

experimental, academic, and industrial investigators would be desirable. Such collaborations, even pairwise, were rare, so the likelihood of receiving exciting proposals by just publishing a solicitation was quite low.

As it happens, the ACMP PM was extremely fortunate. First, the PM was able to overcome a lack of any relevant subject-matter expertise through a great deal of help and education from past and fellow DSO PMs, as well as members of the Defense Sciences Research Council during VIP's technical formulation, team building, inception, launch, and duration. Second, the directors of NSF's Division of Mathematical Sciences and the recently formed Office of Multidisciplinary Activities were actively seeking to take their constituencies in multidisciplinary directions. As a result, DARPA and NSF co-sponsored a portion of the VIP program aimed at drawing more NSF-funded researchers towards problems of interest to DARPA. The high technical standards of VIP were attributable to the technical attention paid to it by multiple DSO and NSF PMs with diverse backgrounds. Finally, a special year in materials science at the NSF-funded Institute for Mathematics and its Applications[31] was taking place the year VIP ramped up, so many mathematical and materials scientists with relevant expertise and interests were already in one place. Two workshops held in Minneapolis were the start of the needed VIP technical exchanges and team building. Even so, laying the groundwork for VIP took two years after VIP was pitched and funded by DARPA management, often in the face of great skepticism and even hostility.

Subsequent thematic ACMP efforts included Robust Uncertainty Management (RUM), Predicting Real Optimized Materials (PROM), Quantum Control, Protein Design Processes (PDP), Enabling Quantification of Uncertainty in Physical Systems (EQUiPS), and Lagrange.

1.1.4 Theme 4: Architecture-aware algorithmic representations

The rate and degree of adoption for novel computer architectures generally depends on the quality of available programming models and automation tools for maximizing performance and programming productivity. By the 1980s, architectures were increasingly complex and diverse, and software tools had not kept pace with hardware advances.

After the divergence of CS and mathematics as disciplines, mathematical research related to computing focused primarily on either theoretical CS or computational science. In both cases, algorithms were routinely developed based on high-level architectural considerations (e.g., sequential, parallel, vector) with little regard to lower-level details distinguishing platforms within the same architectural class. Frequently, details that had outsize impact on actual performance were often ignored or undisclosed by vendors. The algorithm selection and mapping problems for performance-critical situations were addressed through combinations of heuristics, compilers, and human programming expertise. To achieve high performance for critical applications, hand tailoring of algorithms and code was often needed to take advantage of low-level machine considerations. This model of code development is expensive, error-prone, and unsustainable.

Early on, ACMP PMs recognized the great potential of expanding the mathematical scope of CS and computer engineering (CE) research. DSO mathematics projects pioneered novel multidisciplinary CS and CE research that resulted in entirely new software engineering automation methodologies and technologies that accelerate adoption of new classes of computer architectures. ACMP's projects were unique in their cohesive approach to mathematical research that emphasized co-design encompassing mathematical formulations, programming tools, computer architectures, and application needs.

[31] https://ima.umn.edu

This theme's impetus was a 1980s ACMP project that showed how "architecture-aware" group-theoretic representations of FFT algorithms could be the basis for systematized performance tradeoffs (see §2.9). The possibilities offered by architecture-aware representations inspired the Optimized Portable Algorithms Libraries (OPAL), Optimized Portable Algorithms and Applications Libraries (OPAAL),[32] and Design and Exploitation of Structure in Algorithms (DESA) programs that invented and developed such representations for a broad range of algorithms and platforms in signal processing modeling, and simulation. These programs sought to facilitate new collaborations between disparate disciplines and perspectives. Ideas and collaborations initiated in some of these programs' projects gave rise to new projects in other ACMP themes and DARPA offices (notably, MTO and Information Innovation Office [I2O]).

One project, Signal Processing Implementation Research for Adaptable Libraries (SPIRAL), built on prior ACMP FFT projects. SPIRAL innovations caused a paradigm shift in the context of several important classes of FFT-based signal processing by replacing ad hoc approaches with a formal system for solving the mapping problem in a rigorous way that was capable of reasoning like human algorithm developers and expert programmers (see §2.10). Specifically, SPIRAL researchers extended previously developed links between algorithmic mathematical representations and architectural parameters. They also reformulated automatic program generation and performance tuning (for FFT and other domain-specific algorithms as an optimization problem) by combining computer algebra representations with ML search methods. To address the newly emerging workstation systems and multicore architectures, the link between algorithmic mathematical representations and architectural parameters was extended (e.g., to consider SIMD [single instruction/multiple data] vector instruction sets, memory hierarchy, number of cores). This "architecture-aware" representation was used to systematize performance tradeoffs of FFT and other signal processing and linear algebra algorithms, becoming a framework for performance portability.

The most enduring capability found in SPIRAL is its mathematical representational machinery that seamlessly traverses the multiple levels between the algorithm and the hardware. It enables formal knowledge capture of pertinent algorithmic, hardware-related, and program-transformation constructs previously found only in human-understandable form. The system produces results close to or better than hand-tuned codes by combining formal knowledge with high-level reasoning (term rewriting, constraint solving, and backtracking search). The resulting architecture-aware "compiler" concept ultimately inspired high-performance computer vendors like Intel to move from hand-coded to machine-generated scientific libraries for a broad class of related algorithms, resulting in dramatically reduced costs and increased productivity.

SPIRAL's representational capability facilitates "forward compatibility" to future computer platforms and accommodates expression of vast numbers of algorithms. SPIRAL results were later extended to a much broader class of advanced signal/image processing algorithms that includes filtering and wavelets. Efforts continue, on both the forward problem (map algorithms to hardware) and the inverse problem (find hardware specifications for given classes of algorithms or problems), along with work towards chip generation.

The insights underlying this seminal work that applied abstract algebraic representations to computer technology are the result of over 20 years of interdisciplinary research originating with DSO projects. Building the SPIRAL system required mathematicians, computer scientists, and engineers to learn from each other's school of thought to collaboratively reach a synthesized

[32] co-funded and co-managed with NSF

conceptual framework.[33] Without the ACMP PMs' vision and active management over many years, it is doubtful that these technological developments would have occurred.

Often, low-level capabilities of specialized platforms offer broader capability that is difficult to recognize or to realize due to limited programming models or automation tools.[34] ACMP efforts inspired an MTO project[35] focused on real-time space-time adaptive processing (STAP). STAP—unachievable at that time—is a method of cancelling signals that interfere with target detection by using adaptive beamforming and had been of great interest for DoD systems since its invention in the 1970s. As such, it was chosen as an exemplar for a broad class of signal processing computations of interest. Real-time STAP processing requires fast computation involving large matrices. MTO's STAP-BOY project demonstrated that graphics processing units (GPUs) could provide a quantum leap in signal processing capability (see §2.11).[36] GPU architectures were an excellent candidate for accelerating STAP computations because of their parallel structure and small footprint. However, the GPUs' programming infrastructure had been developed for image creation and consisted of graphical image manipulation primitives. The key insight was that an effective mathematical translation between graphical image manipulation primitives and equivalent matrix operations used in STAP would allow advanced signal processing problems to be formulated in a language understood by graphics processors.

Since the computational kernels considered and domain knowledge regarding algorithm and hardware tradeoffs were widely applicable, STAP-BOY's innovations resulted in a multiplier effect in the commercial sector that changed the trajectory of GPU use and technology on a remarkably short timescale: from mainly focused on the niche gaming market to mainstream computing. The GPU market size is expected to reach $157.1 billion by 2022, growing at a CAGR of 35.6%.[37] Moreover, the ubiquity of AI is attributable in part to the ready availability and ease of programming GPUs embedded in modern computing platforms. Use of GPUs in the DoD sector as a standard approach to accelerate advanced image and signal processing in the embedded space is also growing and directly benefits from commercial advances.

The overarching co-design question motivating Theme 4 (and a few projects in Theme 1) was just the tip of the iceberg in terms of the game-changing potential of unexpected math-CS-CE synergies. However, influencing the direction of computing technology and engineering in nontraditional ways requires facilitating novel collaborations that may not yet exist.

1.2 Concluding Observations

As seen, many ACMP projects were at the mathematical forefront in areas where the escalated attention jump-started or accelerated progress towards concrete solutions. For instance, the attention paid to multiscale phenomena across many ACMP projects reflected the need for and the potential of new mathematical approaches in practical contexts such as radar, CEM, and materials processing. Progress in problem areas of interest to DoD was hastened—and, sometimes, caused—by ACMP's enabling of the corresponding, appropriate multidisciplinary collaborations. The influence of the resulting, enduring research communities on the direction of science, technology, and engineering induced interest across a range of academic disciplines and application

[33] M. Puschel, J. M. Moura, J. R. Johnson, D. Padua, M. M. Veloso, B. W. Singer, J. Xiong, F. Franchetti, A. Gacic, Y. Voronenko, and K. Chen, *SPIRAL: Code generation for DSP transforms*, Proc. IEEE, 2005, 93 (2), pp. 232–275.

[34] An early example of this was field-programmable gate arrays.

[35] started by a former ACMP PM and involving investigators from ACMP architecture-aware projects

[36] D. Braunreiter; J. Furtek; H.-W. Chen; D. Healy, *Overview of implementation of DARPA GPU program in SAIC*, Proc. SPIE 6979, Independent Component Analyses, Wavelets, Unsupervised Nano-Biomimetic Sensors, and Neural Networks VI, 19 April 2008.

[37] https://www.alliedmarketresearch.com

domains—evident from the subsequent large increases in the number of new journals[38] and publications on related topics.

ACMP's technical achievements resulted in major advances to the state of the art in many technical fields. Over time, cross-fertilization among these themes and with other domains allowed ACMP to have a broad technical reach. The approach of integrated teams of mathematicians with their science and engineering counterparts beginning in the 1980s was a persistent characteristic of ACMP projects.

Hopefully, this short description of ACMP's initial efforts shows how the work had significant application, industrial, and commercial impact, while also engaging the mathematical community in interdisciplinary problem-solving at the frontiers of technology in areas critical to DoD's mission. ACMP was instrumental in repeatedly translating theory to practical applications tools and to commercial products. Given the conservatism of entrenched communities, this tech transfer only occurred because the funding enabled the necessary fundamental research as well as the requisite experiments and real-world validation. A key to the success across so many diverse projects was involvement by parties that reduced theory to practice and, in several cases, became leaders in the resulting methodologies or technologies.

In conclusion, the author will make a few subjective comments here based on personal observation. Involving mathematicians at the outset greatly streamlines the often circuitous and time-consuming process of developing relevant fundamental mathematics and applying it in practical, high-impact ways. But this is a relative statement, and a first project in a new direction is seldom sufficient to complete this process; rather, it sets collaborations in motion that, hopefully, produce tangible evidence of promising directions to pursue. Since mathematics is general and multi-purpose, there is a likelihood that any breakthrough would have unforeseen implications elsewhere. Therefore, PMs should take care to avoid over- or under-constraining projects. In addition, taking highly speculative mathematical ideas to fruition generally requires more time and much greater resources than are typically associated with mathematics research. To paraphrase the author's response to the DARPA Director when questioned about the (seemingly excessive) level of funding requested: "to get engineering impact, the funding must also be engineering-scale."

While there are vast numbers of intractable mathematical problems of DoD import, ACMP's greatest successes occurred when radical mathematical theories had game-changing implications for applications in need of fresh mathematical approaches or that had never even been considered. Ultimately, ACMP's tack, of taking the long view in pushing mathematical theory forward while advancing the application state of the art when possible, was ideally suited to DARPA's mantra of anticipating the DoD's needs long beforehand and initiating or accelerating progress towards meeting them.

[38] Including numerous journals on wavelets and multiresolution analysis and the SIAM interdisciplinary journal, Multiscale Modeling and Simulation, started in 2003.

2 Supplementary Project Details

2.1 Multiscale and Time-Frequency Methods

ACMP's wavelet-related funding began with some groundbreaking exploratory projects to explore whether wavelet-related mathematics could overcome some longstanding impediments to high fidelity and high throughput in important DoD applications such as radar processing and data compression. Some projects were remarkable in that they involved collaborations between pure mathematicians and DoD domain experts from the outset. ACMP's impact is attributable to the momentum caused by these early projects and the multi-office and ACMP programs that followed.

One exploratory project showed that for navigational purposes, accurate location could be maintained for images that had been highly compressed. Another project demonstrated that for certain representative test cases, wavelets provided superior performance and reduced computational loads in millimeter-wave radar ATR by virtue of their compact signal representations and superior computational scaling compared to FFTs. This latter project was the forerunner of a series of ACMP research projects collectively known as the Longbow insertion demo,[39] that ultimately achieved significant classification performance improvement over a production radar system that was still under active development.

Digital imagery had become prevalent by the early 1990s, but with image depths of 8 to 24 bits per pixel, the resulting file sizes were very large for the storage and communications technologies of that time. The JPEG digital image compression standard had already been developed and come into use, but the visual quality of JPEG-compressed images degraded significantly at higher compression ratios because the algorithm broke the image into 8 x 8 blocks and compressed each of them using a block discrete cosine transform. At high compression ratios, the image would be reduced to an approximate representation made up of constant values across 8 x 8 blocks. As the volume of digital imagery increased, storage and communication bandwidth limitations became significant impediments to routine use of digital imagery. The need for higher-quality small-file-size representations became paramount—particularly for application domains with national security, legal, safety, or health implications.

Wavelet algorithms emerged as a synthesis of multiscale analysis and subband coding (transform-based compression) for signals, beginning in the late 1980s. Through a hierarchical cascade of multichannel filter banks, iteratively decomposing the low frequency results, one arrives at a transform representation that holds information at different scales. When this transform representation is quantized to produce a lossy, compressed result, the information loss is more gracefully spread across the scales of the image. Thus, at intermediate and high compression ratios, a wavelet-compressed image has higher visual quality (measured both objectively via peak signal-to-noise ratio and subjectively via human tests).

A variety of contributors had been working on multiscale analysis and subband coding for some time. However, to make wavelet-based compression suitable for wide use, one needed to be able to choose the specific transform (both the filters involved and the adaptation of the cascaded tree structure to the data) and feed the transform coefficients into an efficient quantizer that would result in a user-specified compression ratio, as opposed to producing an arbitrary file size.

Multiple teams funded by DARPA contributed innovations to overcome these obstacles and create wavelet compression implementations suitable for real-world use. The researchers at Aware, Inc. developed arbitrary tree structures of multiband (not just high/low) wavelets and an iterative

[39] C. Stirman, *Application of Wavelets to Automatic Target Recognition*, Final Report, 1995, https://apps.dtic.mil/sti/citations/ADA294854.

quantization algorithm that delivered predictable results efficiently. Coifman's group (Yale University) developed entropy-based best-basis methods for constructing the tree decompositions. The Aware team also found a computationally efficient implementation for a symmetric (linear-phase) wavelet transform pair identified by Daubechies and Lagarias that produced superior images.

The combination of these innovations resulted in lossy wavelet-based compression that could be tailored to a specific application domain (e.g., tree structures were optimized and selected for the FBI's Wavelet Scalar Quantization [WSQ] standard), delivered scalable compression that would continuously vary image quality as the user dialed in the compression ratio, was fast to compute, and resulted in small file sizes. The wavelet transform's multiscale nature naturally supports progressive decoding.

These techniques have made their way into multiple technical standards for the compression of fingerprints, medical images, and general still images and motion capture. Wavelet compression is widely used today to preserve superior image quality at high (lossy) compression ratios (small file sizes), while offering smooth transitions at intermediate file sizes. Wavelet methods were embodied in the WSQ and JPEG2000 image compression standards (JPEG2000 was incorporated into the Digital Imaging and Communications in Medicine Standard [DICOM]). As a result, they are the standard of use for

- fingerprint compression / biometric identification;
- medical diagnostic imaging, such as MRI, CT, or X-ray scans;
- digital cinema and video broadcast production; and
- wireless multimedia.

2.2 FMM for the Helmholtz and Maxwell's Equations

A perennial concern in the design and deployment of military aircraft is the minimization of radar cross-section. However, despite advances in computing power and software, in the mid-1980s the use of CEM simulation—in contrast to CFD—was still not accepted as a reliable method for gaining detailed knowledge of radar scattering details of a contemplated design. CEM codes needed to be robust enough to cope with a broad range of frequencies and detailed geometric models, while being sufficiently fast and accurate to supplant costly physical experimentation. This was far from true at the time. The most pressing impediment to virtual experimentation was the absence of scalable, accurate algorithms for solving wave equations in regimes relevant to aircraft design.

The dominant technical issue was the lack of fast, accurate iterative solution methods (performing successive approximations until convergence) for solving boundary integral equation formulations of wave equations in regimes involving high-frequency electromagnetic waves. For example, since 1 GHz radar has a 30 cm [~1 ft] wavelength, the aircraft were 10s of wavelengths in both length and width. For a single case, the codes of the day either ran for days/weeks or ground to a halt, whereas the desired turnaround was in hours per run.

Existing iteration methods involved multiplication by dense matrices, which required at least cubic scaling with problem size. Since very large numbers of unknowns were required to resolve the complex geometry of aircraft components—let alone full vehicles—simulation for realistic problems was intractable on any existing or foreseeable computing platform. For example, a problem with 10^6 unknowns yields a dense matrix of dimension 10^{12}. Furthermore, inaccurate methods for describing fine geometric details of the object being considered resulted in slow or non-convergence of the iterations.

The fundamental insight of L. Greengard and V. Rokhlin's seminal work on the FMM[40] was that multipole expansions[41] could be used as an approximation tool and that, combined with sophisticated data structures, could be systematically exploited to reduce the cost of computing all N^2 interactions in a system of N interacting particles from the brute force order-N^2 computations to linear in N. The result was a fast, parallelizable algorithm where the operator is applied in a divide-and-conquer manner using sparse, rather than dense, matrices. With 10 million boundary points, speedups of 3–4 orders of magnitude were achieved compared to a naïve algorithm. While originally designed for electrostatic problems, the high-frequency FMM,[16,42] subsequently developed by Rokhlin, extended this capability to scattering problems, with the discovery that multipole expansions could be transformed to another basis for which shift and translation operators were diagonalized. This led to the first frequency-domain algorithm that scaled nearly linearly with the problem size (requiring $N \log N$ operations).

Aside from "fast algorithms," two further problems needed to be addressed to fundamentally reinvent electromagnetic simulation capabilities. One was the lack of well-conditioned integral equation formulations that were insensitive to mesh refinement and stable at all frequencies. The second was the lack of high-order quadrature methods that could be applied to scatterers of arbitrary shape. These problems were addressed beginning shortly after the development of the FMM in work by several groups.[43,44,45] Quadrature design has a long history and is, even today, a subject of ongoing research. An important insight from the ACMP projects was that direct discretization of the integral equation using a "Nystrom" method would lead to significant acceleration when compared to the standard "Galerkin" methods that were (and are) in widespread use. Galerkin methods yield somewhat better accuracy with low-order discretizations, but the difference becomes negligible at high order. The Galerkin approach, however, requires a second and costly integration step which is avoided in the Nystrom formalism—hence the superior performance once such quadrature rules became available.

2.3 Fast Algorithms for QC and Beyond

In the mid-to-late 1990s, chemistry and materials science impacted many fields: the chemical and petroleum industries, batteries, materials and integrated circuits fabrication, as well as drug design. At that time computational chemistry and materials science simulations were just beginning to yield practical results with enough accuracy to replicate or predicate experimental results for small molecules. Before then, computations were used to confirm results obtained using heuristic knowledge. The goal was to replace some experiments with simulations to improve manufacturing and reduce cost. At the time, a variety of numerical methods (often re-invented) were used. However, these existing methods were often limited in both accuracy and scalability.

ACMP funded several collaborations between mathematicians and material scientists to see if these limitations to large-scale virtual experimentation could be overcome. While some advances resulted from these interactions between mathematicians and materials scientists, real high impact required further research. For the effort described herein, the seed planted led to significant

[40] a technique for rapid application of integral operators governed by electrostatics (the Laplace equation)

[41] invented around the time of Maxwell

[42] V. Rokhlin, *Diagonal forms of translation operators for the Helmholtz equation in three dimensions*, Appl. Comput. Harmon. Anal., 1, Academic Press, San Diego, 1993, pp. 82–93.

[43] H. Contopanagos, B. Dembart, M. Epton, J.J. Ottusch, V. Rokhlin, J.L. Visher, and S.M. Wandzura, *Well-conditioned boundary integral equations for three-dimensional electromagnetic scattering, IEEE Trans. on Antennas and Propagation*, 50 (12), 2002, pp. 1824–1830.

[44] C. L. Epstein and L. Greengard, *Debye sources and the numerical solution of the time harmonic Maxwell equations*. Comm. on Pure and Applied Mathematics, 63 (4), 2010, pp. 413–463.

[45] J. Bremer and Z. Gimbutas, A Nystrom method for weakly singular integral operators on surfaces, J. Comput. Phys., 231, 2012, pp. 4885–4903.

advances, initially in QC, a decade later.

The equations of QC are based on simplifications of a system of multi-particle time-dependent Schrodinger equations and/or density functional theory (DFT). At the time, methods of solving these equations (for finding bound states) used an invented basis set for each problem (using theoretical, computational, and experimental results) to minimize the energy of the system. The problem was that basis-set flaws placed limits on accuracy, the so-called "basis error." Traditionally, the self-consistent solution methodology is limited by solutions to dense eigensystems and Poisson equations, which resulted in at least cubic scaling, limiting the size of problems that could be dealt with, even on parallel computers.

Around 2003, using analysis-based methods, a group of mathematicians headed by Gregory Beylkin (U Colorado) constructed, for a user-specified accuracy, an efficient representation of relevant operators and functions using separated representations and multiwavelets, thereby yielding adaptive bases for computations. This work motivated the subsequent collaboration with George Fann and Robert J. Harrison (both at ORNL at the time),[46] which resulted several years later in the MADNESS software,[20,21] funded by DOE, NSF, and DARPA (HPC). MADNESS solves the QC equations in three or more dimensions using high-performance parallel computers, with computational cost that scales nearly linearly with respect to system size (the number of particles, electrons, protons, …). MADNESS is now considered one of the most accurate approaches in the field. Moreover, the techniques described here are also useful in solving time-dependent integro-PDEs in general.

MADNESS's capabilities include solving equations that arise in Hartree-Fock, DFT, and the Coupled Cluster Method in chemistry (e.g., analytic derivatives, response properties, asymptotically correct potentials, excited state, open systems). These methods were further extended to solve problems in nuclear physics, which are important in the study of exotic nuclei, fission, and fusion. The parallel run-time of MADNESS has been used to implement a wide variety of features, including graph optimization and runtime load balancing, with numerical algorithms scaling beyond millions of cores. From a mathematical perspective, MADNESS emphasizes rigorous numerical precision combined with computational performance and compressed representations of operators and functions—each with its own adaptive representation. Since the accuracy is controlled by the algorithm, scientists do not need to design a grid. Thus, adaptive discretization and data structures of the problems are effectively mapped onto distributed memory on massively parallel computers.

Recently, Beylkin developed a novel approach to adaptive algorithms[47,48] that should eventually compete with and, perhaps, supersede MADNESS. The collaboration with Harrison has continued with recent work on relativistic QC. Besides QC, this research also impacted the development of algorithms for computing with functions of many (100+) variables; some of these algorithms are now commonly used. Further work in this area should affect control theory, data sciences (AI/ML), 3D printing, quantum information/computing, and algorithmic challenges for exascale computers.

---

[46] R. Harrison, G. Fann, T. Yanai and G. Beylkin, *Multiresolution quantum chemistry in multiwavelet bases,* in P.M.A. Sloot et. al. (eds.), Lecture Notes in Computer Science. Computational Science-ICCS 2003, Vol. 2660, Springer, 2003, pp. 103–110.

[47] G. Beylkin, L. Monzon, and X. Yang , Adaptive algorithm for electronic structure calculations using reduction of Gaussian mixtures, Proceedings of the Royal Society A, 475: 2226, 20180901, 2019, https://dx.doi.org/10.1098/rspa.2018.0901.

[48] G. Beylkin, L. Monzon, and X. Yang , *Reduction of multivariate mixtures and its applications*, J. Comput. Phys, 383, 2019, pp. 94–124, https://dx.doi.org/10.1016/j.jcp.2019.01.015.

2.4 Automated Optimal Filter Design

An ACMP project involving wavelet research produced a promising new approach to designing approximations to infinite impulse response (IIR) filters[49] based on factored-polynomial finite impulse response[50] (FIR) filters.[51] IIR filter design is well developed and produces excellent filters but requires a complete signal (from the beginning to the end). In real-time applications (e.g., radar), where the signal is received continuously, IIR filters cannot be used (they are unstable) and, instead, FIR filters must be used. Previously, accurate FIR approximations of IIR filters generally required polynomials of high degree that are inefficient computationally. The new approach provided a means of approximating certain IIRs by factoring the polynomial in such a way that its application in that form as a FIR filter would be inexpensive.

ACMP initiated an OPAL project to determine whether the new approximation method could be applied to filter design for use in a radar digitizer where the very steep filters required were difficult to design by conventional methods. The result of this project was that the new method produced only slightly better designs than those produced using conventional methods. The main advantage was that the factored approximation approach allowed an automatic construction of excellent filters whereas the traditional construction of FIR filters for real-time applications requires an experienced engineer working for (perhaps) several days to come up with an acceptable design. As part of the project, in keeping with OPAL objectives, an automated optimal filter design software tool was developed in collaboration with a small computer engineering design company. While this tool failed to gain traction due to resistance on the part of the signal processing community, nonetheless, a seed was planted.

Now, after considerable subsequent research and SBIR funding by AFOSR, a startup company, The Numericus Group (TNG), has developed software with capabilities[52] well beyond the goals of the original ACMP projects, including

- Automatic design of phase compensation filters (including phase denoising)
- Automatic design of frequency selective filters
- Fast resampling for arbitrary sample rates
- Rapid computation of the cross-ambiguity function (CAF)

These algorithms were incorporated into X-Midas.[53]

Ongoing research has strong implications for a wide variety of new applications. For instance, TNG's automatic filter design methodology should be useful in adaptive control and, additionally, could allow design of device-specific filters that compensate for manufacturing defects. Fast CAF algorithms have been developed that are expected to be enabling in a variety of challenging scenarios, including real-time GPU-based CAF, which must be currently done offline.

2.5 Gravitational Field Calculations

ACMP initiated a project in the late 1990s because of a meeting involving the United States Air

[49] which are rational functions

[50] which are polynomials

[51] G. Beylkin, *On factored FIR approximation of IIR filters*, Appl. Comput. Harmon. Anal., 2, 1995, pp. 293–298.

[52] G. Beylkin, R.D. Lewis, and L. Monzon, *On the design of highly accurate and efficient IIR and FIR filters*, IEEE Trans. Signal Processing, 60 (8), 2012, pp. 4045–4054.

[53] https://wiki.ice-online.com/X-Midas

Force Space Command (USAFSC) that exposed the need for fast, accurate tracking of satellites and/or space junk numbering in the tens of thousands. The issue has been (and still is today) that the model of the gravitational field is based on spherical harmonics, which are global functions. Continuous data gathering on the earth's gravity field, necessary for DoD and other applications, makes the use of spherical harmonics problematic since all the coefficients of the series change even when the data changes only locally. Using this model to evaluate the gravitational force is expensive since the cost grows quadratically as the order and degree (i.e., accuracy) of the model increases. The cost of computing a single orbit must be multiplied by the number of objects in the catalog, making the total cost significant. Most of the cost is in the evaluation of the gravitational force—the problem to be addressed.

The solution[54] used local functions (instead of the spherical harmonics) to represent the gravitational field. While this approach is obvious (and apparently was suggested earlier as well), its technical implementation appeared difficult given the accuracy requirements. The researchers invented an inexpensive solution for constructing a local representation within the required accuracy. Note that the accuracy requirement was very strict: the error in computed position of a satellite after 10 days should be on the order of centimeters in comparison with the spherical harmonic model. While the required accuracy was achieved, note that this requirement is excessive: in a 10-day period, a satellite makes 100–150 orbits, and on each orbit, the actual position is off by (at least) 1 meter compared to the computed position due to various unaccounted-for forces (e.g., residual atmosphere, solar pressure).

The group at USAFSC tested the new model to their satisfaction, but no actions were taken to use it. About 10 years later a student at U Colorado, Brandon Jones, used the model successfully for computing the Moon's gravitational force (during summer employment at NASA Johnson Space Center),[55] which generated interest at AFOSR to start using it at USAFSC. USAFSC's subsequent use of this model, which allows a higher order model to be used requiring lower runtimes, occurred after an AFOSR STTR grant facilitated transfer of the model to Omitron, a company that supports USAFSC.

Notwithstanding, some further comments are in order. In addition to the inefficiencies of updating gravitational field models locally, the spherical harmonics are oscillatory functions and using them to represent a non-oscillatory gravitational field is unreasonable for computational purposes and creates serious technical difficulties from the outset. Alternative representations should be developed and used instead, but the need for alternative gravity estimation tools have not been widely accepted.

2.6 Control of RTP

In the early days (1970s–1980s) of semiconductor processing, most thermal processes consisted of placing multiple wafers in a quartz rack and placing the rack (or boat) into a well-insulated furnace. The wafers were heated to the suitable process temperature and then cooled. Since the rack of wafers was heated from the edge, the temperature ramp rates had to be fairly slow to prevent edge-to-center temperature gradients that could cause slip defects in the silicon wafers. Such furnace technology had many advantages, such as good process uniformity and high thermal efficiency. However, since the furnace was basically a well-insulated tube, the cooling rate was slow.

In the early 90s, transistors became smaller, and shallow junctions were required. The long

[54] G. Beylkin and R. Cramer, *Toward multiresolution estimation and efficient representation of gravitational fields*, Celestial Mech. Dynam. Astronom., 84 (1), 2002, pp. 87–104.

[55] Note that the spherical harmonic gravitational model for the Moon is much more expensive to evaluate than that for the Earth.

thermal cycles of furnaces caused the ion-implanted dopants to diffuse too deeply into the silicon. Thus, furnaces were being developed that were believed capable of using rapid heating of silicon wafers to limit dopant diffusion.

For the previous generations of slow furnaces, the temperature could be controlled with a few independent zones, but in RTP chambers, the entire wafer was heated across its face and many more heating zones needed to be controlled simultaneously. In addition, the multiple zones had a significant amount of crosstalk, with neighboring zones having significant effects on the temperature distribution across the wafer radius. With the transition from 200 mm diameter wafers to 300 mm wafers in the mid- to late 90s, the number of zones increased again, and the multivariable nature of the control problem became even more significant. The control problem was complicated by several factors such as the nonlinear nature of radiated losses scaling as the fourth power of temperature, the nonlinear spectral and temperature dependence of optical properties of wafers and other chamber components, and the need for multi-point control to ensure not just tracking fidelity but also good temperature uniformity across the wafer surface. In addition, the heating/cooling rates needed to be dramatically sped up, with ramp rates approaching 300° C/s over a range of temperatures from 300° C to 1100° C. The then state-of-the-art, simple black-box control schemes were inadequate to the task because of the complexity of the underlying physics. Therefore, the controller design needed to be more mathematically sophisticated while allowing sufficient computational speed for *in-situ* control.

SC Solutions had been building mathematical models of semiconductor thermal processing equipment for many years, mostly to aid in equipment design.[56,57,58] However, at that time (mid-90s), it was believed that mathematical models of the thermal process would require large systems of ODEs to be useful and would be far too slow to compute in real time, so use of physics-based models in real-time controllers had not previously been contemplated.

In this project, mathematical techniques were developed that significantly reduced the order of the high-fidelity physical models while maintaining the accuracy needed for feedback control. The resulting models were also computationally fast so that they could be integrated into the control design to automatically accommodate the nonlinear, radiation-dominated process and dramatically improve the dynamic temperature uniformity (< ±1° C across wafer).[59,60] Few people had believed that a model with a few hundred states could be made sufficiently accurate.

Methods for integrating the multivariable control with the real-time models were initially tested in the late 1990s on a commercial RTP system. For the first time, equipment makers could evaluate the closed-loop performance of RTP equipment in simulation and efficiently evaluate design changes. Figure 1 shows a comparison of the simulated versus actual equipment temperature response and corresponding heater commands. The simulated response using the model is remarkably like the measured actual response in a commercial RTP system for a fast-ramp shallow-junction anneal process.

---

[56] K. F. Jensen, T. P. Merchant, J. V. Cole, J. P. Hebb, K. L. Knutson, and T. G. Mihopoulos, *Advances in rapid thermal and integrated processing*, in Proc. NATO Advanced Study Institute, F. Roozeboom, ed., Kluwer Academic Publishing, Dordrecht, The Netherlands, 1996.

[57] J. Ebert, A. Emami-Naeini, R. L. Kosut, *Thermal modeling of rapid thermal processing systems*, 3rd International RTP Conference, August '95, Amsterdam, Proc. RTP-95 pp. 343–355.

[58] J. L. Ebert, A. Emami-Naeini, H. Aling. and R. L. Kosut. *Thermal modeling and control of rapid thermal processing systems*, in Proc. IEEE Conf. Decison and Control. Dec. 1995.

[59] H. Aling, J. L. Ebert, A. Emami-Naeini, R.L. Kosut, *Application of a nonlinear model reduction method to rapid thermal processing (RTP) reactors*, in Proceedings of IFAC-96.

[60] H. Aling, S. Banerjee, A. K. Bangia, V. Cole. J. Ebert, A. Emami-Naeini, K. F. Jensen, I. G. Kevrekidis, S. Shvartsman, *Nonlinear model reduction for simulation and control of rapid thermal processing systems*, in Proc. American Control Conference, 1997, pp. 2233–2238.

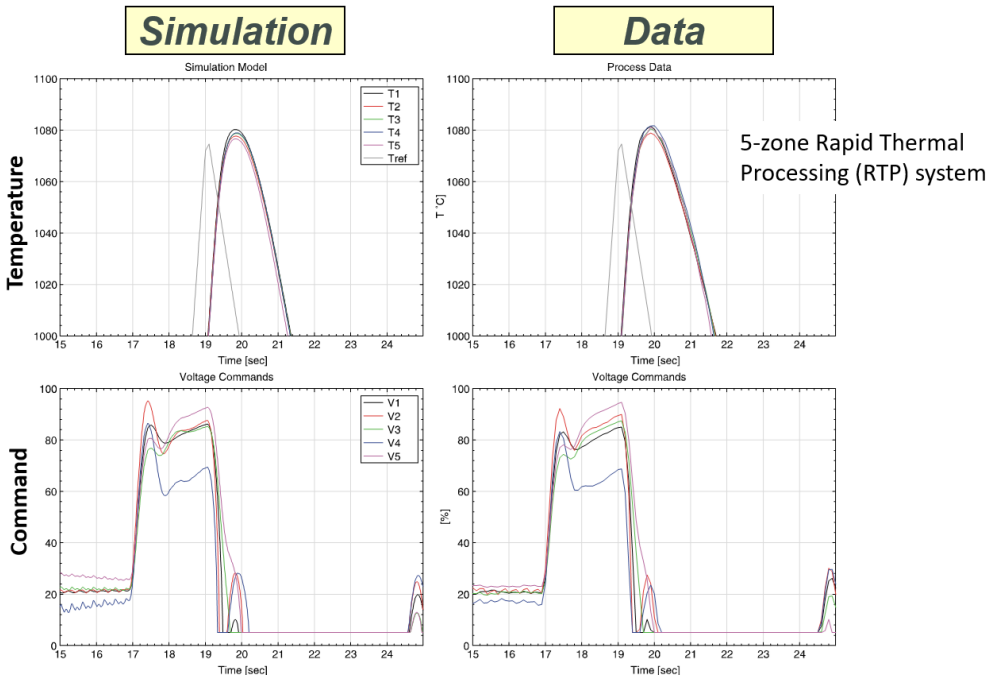

Figure 1. Response comparison

2.7 Simulation of Epitaxial Growth

This project focused on the growth of III-V materials for high-speed electronics, for which HRL (formerly Hughes Research Laboratories) is a leading developer (e.g., of quantum device technology for use in high-bandwidth ultrahigh-speed wireless communication applications). The aim was to develop growth models and in-situ control methodologies for MBE that could be used to impact device performance by improving interface characteristics.

Previous simulation methods were either fully atomistic (e.g., kinetic Monte Carlo [KMC]) or fully continuum (e.g., the phase field method). These methods suffered from the tradeoff between computational speed and physical fidelity. The new island dynamics model developed was multiscale, with atomistic resolution in the growth direction, but continuum in the lateral directions. This model was simulated with novel LSM that described the step edges or island boundaries between regions of different atomistic height, as well as a density for adatoms on the surfaces. Diffusion of adatoms and their attachment to steps or islands was simulated by a continuum diffusion equation, and the stochastic nucleation of new islands was included. This model was validated by comparison to KMC simulations and by data from RHEED and STM imaging.

An example of the simulation results from this project is a demonstration that deterministic or uniformly random nucleation leads to inaccurate results on layer-by-layer growth, but that probabilistic nucleation at a rate proportional to the square of the adatom density (or a higher power, when the critical nucleus is larger than a dimer) corrects those inaccuracies.

The LSM developed in this project for the island dynamics model have become standard methods for simulation of epitaxial growth. Its original application was for layer-by-layer growth. In subsequent work, this methodology was developed further to model multilayer growth (the formation of mounds). It has been combined with elastic models to model the formation of so-called quantum dots. The new methodology was also the starting point for the development of new adaptive multigrid methods.

ACMP was instrumental in bringing together researchers from two distinct fields to work on an important problem that neither could have done independently. LSM were virtually unknown to the materials science community before this program and may well have still been unknown if this program had not happened.

In addition to the significant collaboration between UCLA Mathematics Professor Chris Anderson and HRL Scientist Mark Gyure (who recently moved to UCLA),[61] several industrial postdocs and a mathematics graduate student were influenced by their involvement in the VIP project and have gone on to pursue successful careers in academia, research, program management, and industry.

2.8 Multiscale Modeling and Control of GMR Device Manufacturing Processes

GMR is an electron spin-based effect that was observed in thin film structures fabricated with alternating ferromagnetic and nonmagnetic layers by two European scientific teams independently in the late 1980s.[62] A. Fert and P. Grünberg were awarded the 2007 Nobel Prize in physics for its discovery. It was soon found that the application of a magnetic field to a Fe/Cr multilayer resulted in a significant reduction (typically 10%–80%) in the multilayer's electrical resistance (Figure 2), which prompted scientists worldwide to try to harness the GMR effect's power for sensors and nonvolatile, magnetic data storage.

GMR's biggest application has been in the data storage industry. In addition, its insensitivity to radiation damage also led to great DoD interest for "rad hard" memories. Many people from both industrial and academic institutions have contributed to the advances in hard-disk memory devices over the past 50 years. IBM introduced the first mass storage device based on recording data on hard disks, the Model 350 RAMAC (Random Access Method of Accounting and Control), in 1956. IBM researcher Stuart Parkin's subsequent discovery and application of a "spin valve" (Figure 3)—the ability to change the magnetic state of materials at the atomic level—enabled detection of minute magnetic impulses when flown over a magnetic hard drive, resulting in the ability to write and store vast amounts of data. IBM was first to market GMR-based hard disks in the 1990s.

The discoverers of GMR used MBE in their research. While perfect for R&D, MBE was very complicated and slow in comparison to other deposition techniques and required much more stringent control on impurities and very low pressures, while having a small growth rate.[63] IBM researchers were able to use sputtering techniques instead, which had the advantages of simplicity and cost effectiveness over MBE for industrial-scale GMR device manufacturing. However, high variability in wafer-to-wafer film properties, especially the change in spin valve device resistance as the pair of magnetic layers were switched from the magnetic field being aligned to anti-aligned, was a significant impediment to achieving high manufacturing yields.

The fundamental source of the problem lay with the atomic assembly of the devices. One VIP team (consisting of the University of Virginia [UVA], SC Solutions, Commonwealth Scientific, and NVE Corp.) enabled simulation of the atomic-scale structure of the copper conducting layer and its interfaces with the CoFe alloy layers on either side by using molecular dynamics methods developed at UVA. These simulations showed that during sputter deposition processes, the atoms that condensed on the film had high kinetic energy and intermixed with the previously deposited CoFe

[61] which led to the new numerical method and implementation mentioned in §1.1.3.

[62] P. M. Levy, *Giant magnetoresistance in magnetic layered and granular materials*, Science, 26, May 1992, pp. 972–973.

[63] S. Sharpe, Funding Breakthrough Technology, *Case Summary: Giant Magnetoresistance*, University of Cambridge, available at https://www.cbr.cam.ac.uk/fileadmin/user_upload/centre-for-business-research/downloads/research-projects-output/giant-magnetoresistance-case-study.pdf.

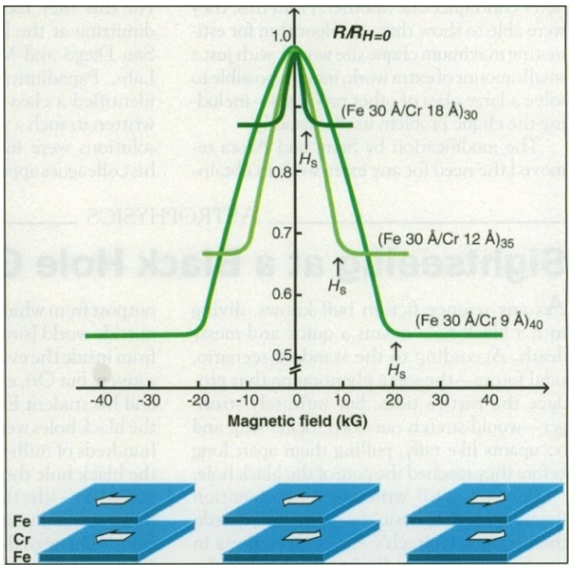

Figure 2: Resistivity vs magnetic field[62]

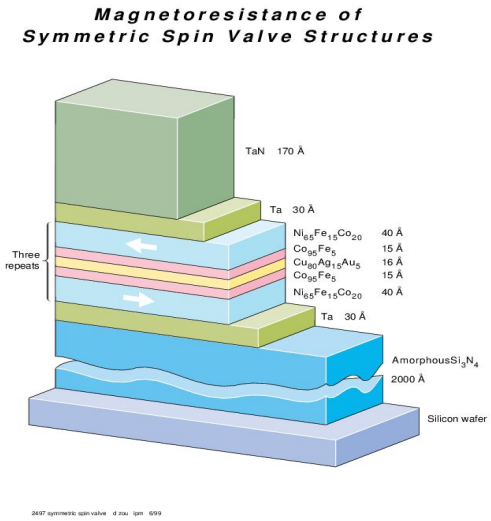

Figure 3: GMR stack

layers. If their energy was reduced, the film interfaces became very rough (wavy), and both phenomena would reduce the change in resistance of the spin valve when its magnetic alignment was switched. The solution was to deposit the atoms with low energy and to then carefully manipulate their assembly to create smooth, unmixed interfaces using low energy inert gas ions (e.g., Ar^+) to move metal atoms across, but not beneath, the growth surface. RF diode sputtering was widely used to deposit thin films on a surface and involves ejection of metallic atoms from a target by energetic Ar^+ ion impacts; their gas-phase transport through the plasma and their assembly on a surface could be electrically biased with a negative potential to control the acceleration of the atomic assembly-assisting argon ions. Real-time control of the underlying atomistic deposition process would be needed to improve the resulting deposition rates and materials quality. In fact, quantitative multiscale models bridging the microscopic-to-macroscopic scales in materials processing generally were nonexistent because methods for such modeling had not yet been developed. Since no model of the entire process from atomistic deposition to process control was available, this VIP team set out to develop one and produced the first instance of an integrated model linking ab initio, gas flow, materials processing, and process control that was used successfully to design and produce an actual implementation on a real process.

This VIP project successfully traversed the long path from the relevant materials physics theory to the implementation of a controller that provided acceptable uniformity via sputtering. The team developed a set of computer models—suitable for process control—for the physical processes that occur in RF diode sputtering for GMR thin film deposition. *Ab initio* models[64,65,66] were developed and linked with multiscale physical models of the actual RF diode sputtering process being used. The associated computational model consisted of three modules that simulated the

[64] X. W. Zhou, H. N. G. Wadley, *Atomistic simulations of the vapor deposition of Ni/Cu/Ni multilayers: The effects of adatom incident energy*, J. Appl. Phys., 84 (4), 1998, pp. 2301–231.

[65] X. W. Zhou, H. N. G. Wadley, R. A. Johnson, D. J. Larson, N Tabat, A Cerezo, et al, *Atomic scale structure of sputtered metal multilayers*, Acta Materialia, 49 (19), 2001, pp. 4005–401.

[66] X. W. Zhou, H. N. G. Wadley, J. S. Filhol, M. N. Neurock, *Modified charge transfer–embedded atom method potential for metal/metal oxide systems*, Phys. Rev. B, 69 (3), 2004, 035402.

various physical phenomena occurring during thin film deposition in an RF sputtering chamber: (a) fluid flow, (b) RF plasma and sputter, and (c) DSMC (Direct Simulation Monte Carlo) transport. Within-wafer thickness uniformity was substantially improved by adopting equipment modifications suggested by the simulations (e.g., target shaping). The team developed model-based control systems for optimal control of RF diode sputtering systems for fabricating GMR devices. Atomistic-level simulations were assembled in reactor-scale models of flux generation and transport so that the effects of reactor design variables and operating conditions on film thickness and atomic-scale structure of GMR spin valve devices could be quantified. In particular, the models quantified the sensitivity of deposition rate and film thickness uniformity to process parameters such as RF power, carrier gas pressure and temperature, and electrode spacing. Hence, the team was able to derive a model suitable for uniformity control and determine the operating tolerances needed to meet the tight deposition thickness specifications.

An integrated target bias-voltage controller was designed and implemented on process equipment at NVE, resulting in significant wafer-to-wafer performance improvement (Figure 4). The plots show the three performance variables (average of measurements at 13 different locations on each wafer), without control (left column) and with time-integrated voltage control (right column). The results of the physical model provided guidelines for selecting process parameters and identifying causes for wafer-to-wafer variability in film properties. This variability was reduced by more than 50% using SC's model-based controller. Specifically, the standard deviation in the average GMR from wafer to wafer was reduced by 65%. The standard deviation in the sheet resistance was reduced by 52%.

In contrast to the Model 350 RAMAC that held 5 MB of data and could barely fit through a 36-inch-wide door, modern disk-drive heads have 1–15 TB storage capacity and fit in very compact packages. An economic measure of the progress in the field is that while the cost of the RAMAC data was about $10K per megabyte, that of a modern drive is less than 1¢ per megabyte.

The critical performance improvements that resulted enabled a quantum leap in GMR use in commercial products used by the DoD. UVA subsequently used the VIP methodology to design a different approach to thin film multilayer assembly in which different ions were used to create the metal atom vapor fluxes and control their assembly on a wafer. This enabled the use of very low energy (5–20 eV) glancing angle Ar^+ ions to optimally configure the interface structures. A DARPA-funded collaboration between 4Wave and UVA then fabricated a BTIBD approach thereby reducing the concept to practice. This tool was put into service at UVA, and a second-generation tool using the same technology has been commercialized to produce ultra-smooth metal and oxide films for both magnetic and superconducting tunnel junctions.

This DARPA research effort was also very complementary to *spintronics*, a nanotechnology also pioneered by DARPA, that relies on electron spin rather than electron charge to acquire, store, and transmit information. NVE, a leader in the practical commercialization of spintronics,[67] had hardware R&D funding from Federal agencies including DARPA. However, its significant commercial penetration (on a very short time scale for new materials) was attributable to ACMP's funding of the needed multidisciplinary research (encompassing materials science, computational science, engineering, and materials manufacturing) to achieve the requisite processing yields and throughput.

[67] i.e., developing a new generation of computers that will store information in quantum bits, exploiting electron spin and other quantum properties

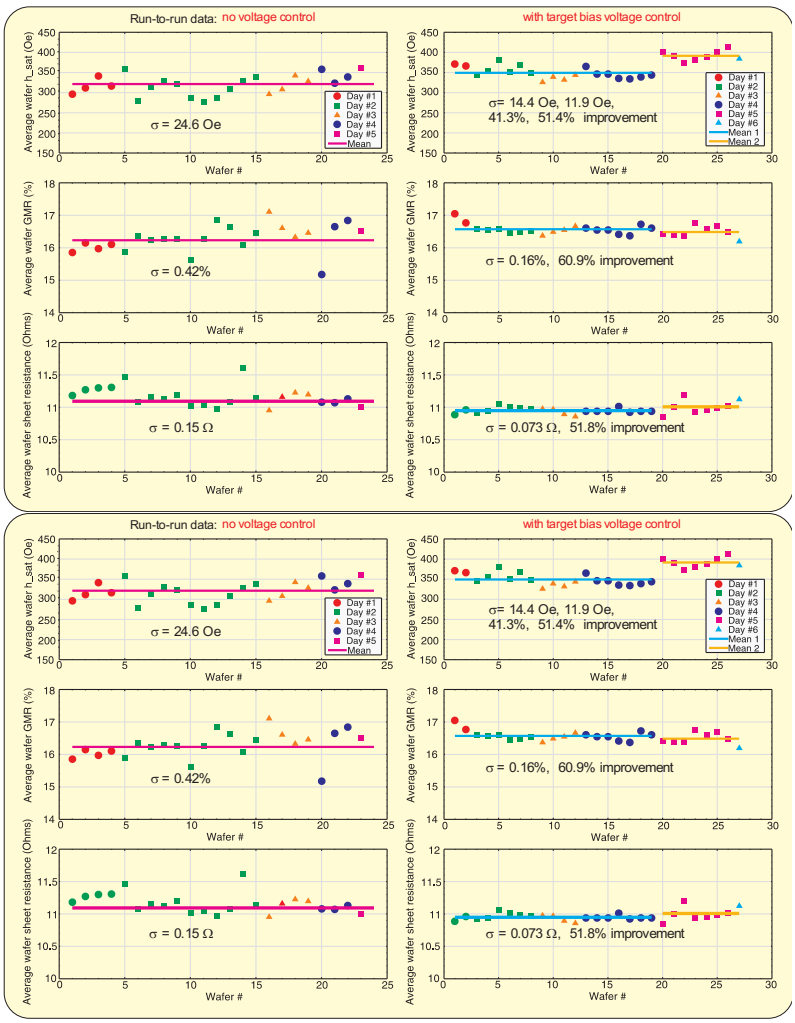

Figure 4: GMR process results without closed-loop control (left column) and with closed-loop control (right column)

2.9 Tensor Product Representations of FFT Algorithms

Fortran's history is an instructive object lesson on overcoming the daunting challenges of programming new computing platforms. Simply put, the driving idea behind the original Fortran compiler was as follows: Write programs for mathematical operations that are as concise as textbook presentations of the respective algorithms. Then, have automatic systems (e.g., compilers) on different computer architectures translate this algorithm description into efficient code. Programmers needed to be convinced to adopt Fortran instead of writing assembly code, and only (almost) no loss of efficiency would convince them. The goal was achieved, and Fortran was adopted.

Many decades later, computer platforms were much more complex, and computing was again in the pre-Fortran dilemma, with programmers writing the most important numerical kernels in assembly. It was not for lack of effort: The compiler community developed optimizing, parallelizing, and vectorizing compilers and had success in many cases; however, matching the performance achieved by human experts when optimizing mathematical kernels proved beyond the reach of compiler systems. They hit a fundamental barrier: human experts' understanding of mathematical algorithms and their interactions allows them to transform algorithms and their implementations in ways beyond the reach of compilers. The key deficiencies in automatic

systems were insufficient capture of domain knowledge and an inability to exploit that knowledge as well as human experts could.

For DoD-relevant applications, no better case in point exists than that of the ubiquitous and important FFT. The archetypal ACMP projects addressed performance portability of FFT algorithms.[68] Historically, many millions of dollars and man hours had been expended on DoD hardware- and software-instantiated FFT algorithms to meet demanding timing and sizing requirements.

Since 1968,[69] it had been known that Kronecker (tensor) products, a well-studied area of abstract algebra, were useful for formulating and studying FFTs. The observation that a certain FFT formulation in terms of tensor products lent itself to parallel processing led to numerous other variants. Some early ACMP projects (mid-1980s–mid-1990s), involving teams of mathematicians and computer scientists, pursued the CS implications of these developments. The researchers cataloged and represented several classical parallel and vector FFT algorithms in this formalism, and the shape of the underlying constructs allowed reasoning about matching algorithms to parallel (supercomputing) platforms of the era. They developed a mapping between certain key algorithmic kernels (permutations and tensor product constructs) and architectural features (mainly vector length and number of processors) that enabled accurate theoretical efficiency comparisons for algorithms derived by certain types of algebraic manipulations.

The effectiveness of this methodology was demonstrated by actual (manual) implementations and generated considerable mathematical interest. By 1998, the methodology developed was well-understood and captured in seminal books.[70] While the idea of automatic numerical software generation for ever-changing new hardware platforms was still a pipe dream, these foundational projects had pointed the way forward.

2.10 The SPIRAL System

The biggest technical impediment faced by the SPIRAL project (late 1990s–early 2000s) was that the algorithms in question were expressed in a form aimed at human consumption rather than automation. Further, optimizing compilers had limited optimization capability with respect to *parallelization* of algorithms with complicated data flows such as the FFT. The advent of increasingly difficult-to-program shared-memory multiprocessors and cache memory hierarchies made sole reliance on human programmers and optimizing compilers increasingly untenable.

To achieve sustainable long-term impact, a fundamentally new approach was needed. Over many years, the SPIRAL project overcame these technical impediments by reformulating natural-language algorithmic information into machine-understandable form and developing an automatic system capable of reasoning like human algorithm developers and expert programmers. To do so, they developed a domain-specific language (DSL) stack that abstracted various aspects of algorithm capture/implementation and enabled full end-to-end automation of performance tuning and program generation, with correctness guarantees.

The first SPIRAL innovation was the development of a new formalism to fully capture knowledge

[68] The FFT discussed here dates to Gauss and was rediscovered and popularized by Cooley and Tukey in the 1960s. It has been described in the 1990s as "the most important numerical algorithm of our lifetime" and was included in *Top 10 Algorithms of 20th Century*, IEEE magazine, Computing in Science & Engineering.

[69] D. Bailey, T. Stockham, M. Pease, …

[70] By R. Tolimieri and by C. Van Loan

required in the automation process embodied in a family of mathematical DSLs.[71] SPIRAL formulates the problem of automatic tuning and generation of algorithm implementations as an optimization problem over the space of alternative representations of the algorithm obtained from a set of constructs combined through breakdown rules. Then a ML search determines the "best" match to the computing platform. Additionally, the DSLs encapsulated structural aspects of the target hardware together with the well-known, as well as novel, program transformations underpinning the ability to tailor algorithms to a wide range of hardware platforms. The SPIRAL system repurposed and significantly extended an existing computer algebra system,[72] tailoring it for matrix algebra to provide the extreme flexibility needed for forward compatibility.

SPIRAL's second major innovation was finding a sequence of program transformations of a given program to a program optimized for an advanced target architecture beyond the reach of compiler technology at the time. These transformations were then used to synthesize correct programs from the algorithm's mathematical semantics, using a rule-based knowledge base and constraint solving via ML search.

Finally, SPIRAL developed an approach to prove correctness of the generated code by 1) ensuring that all representations and rewrite rules had compatible mathematical semantics and 2) showing mathematical equivalence between initial specification and final program via a sequence of semantics-preserving rewrites.

SPIRAL's mathematical representational machinery for a broad class of FFT-like computations was later extended to filtering, wavelets, and more advanced signal/image processing algorithms. SpiralGen, Inc., founded in 2009, was the vehicle for inserting technology in various commercial products (Intel MKL/IPP, Mercury Computing SAL).

Ten years into the development of SPIRAL, support for nonlinear algorithms with FFT-like data flow graphs (e.g., Viterbi decoders, sorting networks, and matrix-matrix multiplication) were added. This generalization eventually unified the space of map/reduce/reshape-style algorithms on high-dimensional data cubes and is related to the polyhedral model in compilation. About 15 years into SPIRAL's evolution, graph algorithms viewed as sparse matrices over semirings, control algorithms, and statistical tests were added. Currently, algorithms in numerous fields (e.g., AI/ML, advanced signal processing, physics simulation, computational science, engineering) are being formalized in the SPIRAL framework. In addition, SPIRAL is addressing integer factorization/modulo arithmetic in the context of post-quantum lattice cryptography and the mapping of quantum algorithms to current quantum computers.

Now, 20 years since inception, SPIRAL's approach and the multidisciplinary community pursuing such approaches has expanded substantially. Code generation for algorithms with clear mathematical semantics across a wide range of parallel platforms continues to be addressed. Formal correctness of the implementations can be automatically verified. Forward compatibility of (mainly data independent) computational algorithms has been achieved (i.e., true "write once, run anywhere"). The SPIRAL system catalogs mathematical algorithms, hardware, and their interactions in a machine-readable form, suitable for high-level reasoning about their interactions. The system provides a math-library-based DSL to effectively capture semantics of algorithms and allows cross-domain and cross-library optimization. SPIRAL efforts continue to address both the forward and inverse problems and are working towards chip generation.

With respect to DoD applications, SPIRAL allows rapid retargeting of core intelligence,

[71] tSPL, SPL/OL, and Sigma-SPL

[72] Groups, Algorithms and Programming (GAP)

surveillance, and reconnaissance algorithms (SAR, STAP, graph algorithms, AI/ML) to novel platforms (e.g., those with constraints on size, weight, power, cost) and achieves performance close to or better than hand-tuned, with correctness guarantees. SPIRAL is also used as a tool for performance engineering and by DARPA contractors (HRL, Boeing).

SPIRAL's impact on software infrastructure is widespread and includes

- online hardware and software generators for FFTs, Viterbi decoders, etc.;
- open-source SPIRAL code and tutorials;
- library front ends: FFTX, GBTLX, NTTX;
- SPIRAL-generated code in industry standard mathematics libraries: Intel MKL and IPP, Mercury Computing SAL, FFTW for BlueGene L/P/Q. FFT library for Cell BE. FFTE kernel generator for #1 supercomputer Fugaku and future IBM POWER 10; and
- becoming the new standard FFT library[73] in the exascale world and beyond.

SPIRAL has benefitted from and broken significant new ground in each of the disciplines involved:

- **Math:** Produced computer languages for describing algorithms in a natural mathematical notation that can be translated to high-performance code.
- **CS:** Influenced a paradigm shift in compilers towards use of search in program transformation selection and automatic performance tuning. Promoted the use of DSLs to generate high-performance code. Pioneered the use of natural mathematical specifications and transformations in verifying the correctness of the generated code and in understanding the algorithm space.
- **CE:** Developed groundbreaking theory and practical application to solving forward and inverse mapping problems, HW/SW co-design, application-specific accelerators, and new algorithms for special-purpose hardware.
- **Signal processing:** Pioneered "algebraic signal processing," a "meta"-SPIRAL—showing that many linear transforms are Fourier transforms for an appropriate signal model and how to derive many existing and new fast algorithms from basic principles—to broaden SPIRAL beyond the FFT. Algebraic signal processing has been extended to process data arising in many new application domains leading to a new theory, "graph signal processing."

The magnitude of the problem continues to grow due to the ongoing, diverse changes in computer architectures. Indeed, it took several ACMP projects to start the process, create momentum, and facilitate the required multidisciplinary research. The approach is still being actively developed and refined, with new technological frontiers continually emerging out of the research.

The impact of this project is attributable to the successful synthesis of the diverse perspectives contributed by mathematicians, computer scientists, and engineers: Mathematicians tease out the structures by which algorithms and computer systems can be represented to enable automation. Computer scientists concern themselves with how automation is achieved including guarantees of performance, applicability, and correctness. Finally, engineers determine what is relevant to capture and model about ever-evolving platforms and what parameter space of algorithms is relevant.

[73] replacing Fastest Fourier Transform in the West (FFTW)

2.11 STAP-BOY: Transformation of GPUs from Niche to Commodity

In the mid-2000s, a sizable gap existed between the needed processing throughput and what was achievable on existing or envisioned platforms for critical DoD signal processing computations. Conventional processor architectures were limited by processor memory bandwidth restrictions and highly serial operation, even when configured for parallel processing. Specialized processors could be designed, but their long-term viability was uncertain because of their limited applicability, high production costs, and lack of chip foundries. Contributing factors to this gap included increasingly challenging scenarios, better algorithms, increases in computational throughput requirements, higher sensor resolution (more data), and reductions in power budgets.

Highly specialized GPUs had been engineered to meet the processing demands of commercial gaming applications. Researchers for MTO's STAP-BOY project recognized that while the existing GPU programming model was not conducive to numerical computation, the underlying COTS architectures were ideal for a broad class of DoD signal processing computations. STAP, a method to cancel signals that interfere with target detection by using adaptive beamforming, had been of great interest for DoD systems since its invention in the 1970s. Real-time STAP—unachievable at that time because it requires fast computation involving large matrices—was chosen as an exemplar for a broad class of signal processing computations of interest.

The GPU architectures' parallel structure and small footprint made them excellent candidates for accelerating STAP computations. The academic-industrial collaboration, involving mathematicians, signal processing experts, engineers, and software and hardware developers, resulted in a programming model and a new generation of graphics processing languages (GPLs) that helped facilitate the broad adoption of GPUs as general-purpose co-processors on today's computers and as the computational backbone for ubiquitous ML and AI.

The three main technical impediments were the representation of image manipulation primitives as matrices, the need for new algorithmic formulations that could take advantage of the parallelism offered by GPUs, and the limitations of existing programming languages in accessing underlying hardware.

The GPUs' programming infrastructure had been developed for image creation and consisted of graphical image manipulation primitives. OpenGL, the GPL used at that time, contained many geometrical transforms applied to images. The key insight was that an effective mathematical translation between graphical image manipulation primitives and equivalent matrix operations used in STAP would allow advanced signal processing problems to be formulated in a language understood by graphics processors.

STAP-BOY developed a GPU-aware formulation of a key STAP matrix computation and the mathematical mapping between core matrix operations and graphical image manipulation constructs, which were used to demonstrate the GPUs' viability for meeting STAP's computational throughput and power requirements. Other applications such as 3D SAR image formation and 3D LiDAR formation and exploitation could then be treated in an analogous fashion.

University mathematics researchers provided expertise on mathematical transformations that would aid in the manipulation of signal processing constructs to GPLs. These same universities were involved in a Theme 4 ACMP project that originated the use of mathematical representations as the basis of structure-specific programming languages and compilers that bridged the algorithm-architecture divide.

A key STAP computational bottleneck was the need to repeatedly perform adaptive QR factorizations for large matrices. Adaptive QR factorization was typically achieved in STAP

implementations by using "rank-1 updating," an inexpensive method for deriving the current factorization from the previous factorization. Rank-1 updating provides orders of magnitude reduction in computational complexity compared to a full QR factorization but is highly serial in nature. The development of a fast "block" QR updating scheme adapted to the parallel structure of GPUs[74] resulted in the ability to achieve highly efficient GPU use.

The success in speeding up STAP processing[75] influenced a migration to GPLs that exposed more of the underlying GPU architecture and attracted broad interest (and some support) from commercial companies such as NVIDIA, Sun Computers, Quantum 3D, Intel, and AMD. Substantial engagement with hardware and software developers for the emerging general-purpose GPU (GPGPU) programming community resulted in collaborations with PeakStream (acquired by Google) and Jacket (acquired by MathWorks), both of which developed compilation tools for GPU-based algorithms, as well as AMD, who was making its way into the GPGPU marketplace. Because of these developments, both performance speed and human productivity were improved by about two orders of magnitude over the course of the project.

[74] D. Healy, D. Braunreiter, J. Furtek, N. Davis, and X. Sun, DARPA STAP-BOY: *Fast Hybrid QR-Cholesky Factorization and Tuning Techniques for STAP Algorithm Implementation on GPU Architectures*, High Performance and Embedded Computing (HPEC) Workshop 18–20 Sept. 2007, MIT-LL.

[75] M. Roeder, J. Furtek, N. Davis, C. Tebcherani, M. Tanida, and D. Braunreiter, Power Consumption of Desktop and Mobile GPUs for IRSTAP Applications, High Performance Embedded Computing (HPEC) Workshop, 23–25 Sept. 2008.

3 A Path Forward: Systems at Scale

ACMP projects amply demonstrated the potential of novel collaborations involving mathematicians, computer scientists, engineers, and application domain experts to change the direction of analysis, design, and implementation of future systems. Several ACMP activities discussed in previous sections were in fact conceived as initial steps at advancing the state of the art in systems design and performance and give some hints as to what may be possible—namely, a fundamental reexamination of systems engineering more broadly.

This discussion is slanted towards holistic approaches to systems-level problems, which are nearly impossible to address piecemeal and have received little mathematical attention. While the discussion uses human-engineered systems as exemplars, analogous challenges and opportunities pervade other important classes of complex systems (e.g., in biology, medicine).

An important advantage of ACMP-type approaches was the mindset of framing research questions based on real-world problem considerations from the outset, which allows more objective and expeditious identification of both mathematical needs and dead ends. For system-level problems, the needed breakthroughs could encompass different scales spanning multiple levels. Three broad classes of futuristic possibilities that might plausibly result from radical advances are discussed without any attempt to assess the existence of mathematical opportunity, which must be left to interested parties.

3.1 Cross-representational and Cross-disciplinary Modeling and Simulation

Several ACMP projects focused on elimination of accuracy and scaling limitations to enable orders-of-magnitude improvement sufficient for full-scale modeling and reduction of the amount of experimental validation required. The demand for this type of algorithmic breakthrough is omnipresent, even in mature fields such as CFD. In many applications, however, overcoming the immediate computational roadblocks is only a necessary, but insufficient, step in dealing with realistic full-system problems.

As a case in point, FMM was only the beginning for CEM. While FMM eliminated the existing bottlenecks, existing codes are far from able to deal with the scale and complexity of many DoD systems. For complete airframes, designs for different sections are handled by separate engineering teams and subsequently joined to form the actual surface used in the FMM calculation. Conventional “join” algorithms lead to inaccurate surfaces that also have gaps and unintended intersections. Often, the time for human interaction with the “joined” surface to correct errors, rather than the FMM itself, is the rate-limiting step for the calculations. This obviously gets worse with scale (e.g., ships) and remains a problem whose solution requires out-of-the-box, non-incremental approaches.

Among the many additional challenges facing designers of advanced ships and aircraft are issues arising from improving command and control through more sophisticated communication systems. As the number of hardware components grows, the interactions between communications antennas, weapons systems, other EM sources, and the ship/aircraft itself become increasingly important. Such systems have long since reached a stage of complexity that design by experimentation (even with scale models) is infeasible.

Thus, to deploy functioning systems, avoid electromagnetic interference, or minimize radar signatures, accurate computational approaches and new simulation environments are required to address the complexities of realistic environments. These will require all the components mentioned above. In addition, a final requirement for the largest scale problems is the creation of

solvers that go well beyond FMM-based methods to circumvent excessive iterations caused by proximity to true physical resonances.

But electromagnetic considerations are not the only basis for design decisions. Notably, CFD considerations are paramount in high-performance vehicles of various kinds. Therefore, efficient 3D geometry representations for cross-disciplinary use (e.g., in CEM, CFD) are imperative. Translating between different representations, as is done currently, introduces errors, and the impact of these translations in what amounts to a nonlinear calculation is unknown.

3.2 Write Once and Run Anywhere

While major steps were taken in projects such as SPIRAL to develop forward methods for mapping software to hardware and towards inverse methods, this work pertained to a specialized, albeit extensive, portfolio of algorithm classes. There are many open questions as to the extent to which architecture-aware theory and approaches can be generalized. Initial developments suggest that considerable advances in theory and practice are possible, and a whole community well beyond SPIRAL has sprung up that is building on the core insights, making the research area ripe for the next quantum leap. As an example, the evolution of compilers into SPIRAL's architecture-aware technologies suggest that STAP-BOY's impressive impact on compiler technology could be an early step in automatic "write once and run anywhere" methodologies for computations with STAP-like primitives.

As a starting point, breakthroughs for the large-scale distributed computing resources made possible by systems-on-a-chip will require mathematical representations and methodology well beyond the current state of the art. In addition to technical impediments, one of the hard lessons learned repeatedly in computing is that it is nearly impossible for tools that require users to understand the underlying mechanics (in this case the mathematical formalisms) to gain traction in the user community.

In addition to the ground yet to be covered in signal/image processing, linear algebra, and graph theory on the one hand and new architectures on the other, there are other important classes of computations that suffer from the same unsustainably high cost and manpower requirements as those discussed here. One notable example is large-scale simulation codes that are continually maintained and ported to bigger machines by human beings to achieve high performance. Many components of such codes are structured in ways that might lend themselves to architecture-aware approaches. Others occur sufficiently often that they warrant deeper consideration. In fact, one could speculate that with a rich enough palette of mathematical theories, it might be possible to automatically port software to new platforms in cases that currently require significant investments of human effort.

3.3 Systems-of-Systems Engineering

Systems of systems arise frequently in DoD-relevant applications and are often integrated hardware/software/sensor/human multi-component monoliths with fragilities that increase with scale and complexity. Over time, the efficacy of such systems degrades for a variety of reasons, including loss of corporate memory, usage in contexts unforeseen by the original designers, and technology "upgrades" that result in unexpected and unintended consequences. Furthermore, lack of understanding of the latter greatly impedes graceful evolution of systems as new technologies appear. The architectures of such systems have significant structure and should therefore be amenable to rigorous mathematical analysis. Any resulting formal tools for reasoning about such systems would be game changing in terms of both legacy systems and design of new systems.

STAP-BOY was but one recent example of a technology that has already changed the course of

commercial computer system development. But while it might appear to be advantageous to use such commercial technologies in DoD systems, adoption has been slow because requirements for DoD systems are unforgiving and involve small numbers of end products compared to the commercial marketplace. Commercial chip technology increasingly employs diverse, complex system-on-a-chip processing architectures that offer potentially revolutionary capability in terms of low-cost deployment of large-scale resources for sensing, weaponry, etc. Harnessing or countering such capability would be distinctly nontrivial.

Based on experience with systems involving natural phenomena, it is plausible that in addition to write-once-and-run-anywhere-type tools, a multiscale modeling, simulation, and control capability well beyond what is possible today would be a natural part of the ability to flexibly and agilely employ new technologies. In this case, at a minimum, the processes would be interactions among components of the system at different scales, where algorithmic and hardware behavior would be micro/mesoscopic and overall system performance macroscopic. Context-appropriate concepts of "control" would need to be developed to ensure that systems stay within expected operating bounds; the systems-oriented model-based control perspective is portable and may be immensely valuable in driving development of a science of systems engineering. There would be untold benefits for systems of systems common to DoD and elsewhere if such capabilities could be developed. A long research agenda would undoubtedly be required, with many nonincremental developments that would only be understood and unfold over time.