Research Summary

July 27, 2014

Prof. Steven Nowick
Department of Computer Science (and by courtesy, Electrical Engineering)
Columbia University, New York, NY

OVERVIEW

My main research is on asynchronous and mixed-timing digital design. Asynchronous circuits have no centralized or global clock. Instead, they are distributed hardware systems where multiple components coordinate and synchronize at their own rate on communication channels. As chips grow increasing larger and faster, power and design-time requirements become more aggressive, and timing variability becomes a critical factor, there are increasing challenges in assembling centrally-controlled synchronous systems. My key goal is to make asynchronous digital design a viable option.

Asynchronous design has the potential to offer significant improvements in performance, energy, reliability and scalability, since it eliminates the rigidity and overhead of the fixed-rate clock, and allow flexible and distributed assembly and communication of components. In particular, it can provide *low power* (components activated only on-demand, without the need to instrument clock gating, and entirely eliminating the global clock); *high performance* (some asynchronous systems have significantly lower latency and increased average throughput, rather than be bound to a worst-case clock rate); *great robustness to timing variability and unpredictability*; and *modularity and composability*. There is also a recent surge of interest in industry in hybrid designs, which connect standard synchronous components (e.g. processors, memories) through flexible asynchronous interconnection networks, forming *globally-asynchronous locally-synchronous (GALS) systems*, where the asynchronous network provides a scalable and reliable integration medium.

However, asynchronous design poses several unique challenges: circuits must be hazard-free (i.e. avoiding the potential for glitches); developing correct and effective computer-aided design tools and optimization techniques has been difficult; and the analysis and verification of such concurrent systems poses particular obstacles**.** These challenges have been major impediments to industrial acceptance. My key goal is to make asynchronous and GALS digital design a practical foundation for hardware system design.

There are currently eight main projects in my group. **Items (i)-(iii) cover some key application areas**, **while items (iv)-(viii) are on developing foundations and infrastructure.**

(i) developing low-power and high-performance **on-chip interconnection networks,** for both high-performance parallel computers (suitable for *"big data")* and moderate-performance low-power embedded systems for consumer electronics, which efficiently and easily integrate heterogeneous components into a single system, and provide significant cost benefits over leading synchronous designs

(ii) developing a promising new class of signal processors**,** called **continuous-time digital signal processors (CT-DSPs),** whose operation adapts dynamically to the input sample rate, and which can accommodate a wide variety of input formats and sample rates with no design change, and which provides a signal-to-error ratio for some applications

which exceeds that of clocked systems, *for use with a wide range of speech and audio applications*

(iii) developing **ultra-low energy and reliable digital design,** using sub- and near-threshold circuits, to support critical applications requiring long battery lifetime and ultra-reliable operation, such as *bio-medical implants (e.g. cancer monitoring), sensors for environmental and infrastructure monitoring,* and *space applications*

(iv) developing practical **high-speed pipelines,** to support high-performance systems

(v) designing robust **mixed-timing interface circuits,** to flexibly accommodate mixed-clock and clocked/asynchronous timing domains

(vi) developing robust **channel encoding for global asynchronous communication,** to support reliable and low-power systems

(vii) developing **computer-aided software design (CAD) tools and optimization algorithms,** for digital system design in each of the above areas, including techniques to perform design-space exploration on tradeoffs between hardware complexity, power, performance and reliability

(viii) promoting **technology transfer** of my asynchronous CAD tools and circuit design styles to industry

The following is a summary of some highlights of my research since the year 2000 (when I received tenure). In addition to the research overview, it also includes a summary of recent grants.

1 ON-CHIP INTERCONNECTION NETWORKS

The goal of this research is to support the design and optimization of high-throughput, low latency, flexible and low-power digital interconnection networks (i.e. *networks-on-chip [NoC's]*) using asynchronous design. Overall, this work aims to demonstrate that asynchronous and GALS NoC's can provide significant advantages in system power, area, latency, and support for heterogeneous interfaces, while still maintaining high throughput – in direct comparison to leading synchronous designs in identical technology. The work is currently funded by 2 NSF grants.

This work is divided into two sub-projects:

(i) Targeting Desktop Supercomputers (joint with University of Maryland, Prof. Vishkin's group)
This work (2008-2013) is in collaboration with a parallel architecture group at the University of Maryland, which is developing the synchronous shared-memory processor environment (called XMT, a PRAM-based massively-parallel architecture) and simulation tools. My group is leading the design of the new asynchronous network. The target topology is a variant Mesh-of-Trees (MoT), which has been shown useful for shared cache access in fine-grained highly-parallel systems. The goal is a *GALS network*, that can integrate multiple synchronous cores and caches operating at unrelated clock rates. Most recent GALS systems have focused on low- to medium-performance embedded applications, or involve advanced circuit techniques for high-performance systems (e.g. low-swing, dynamic logic, wave pipelining, pulse mode). In contrast, this proposal is aimed at medium-to-high end multi-processors, yet uses robust handshaking protocols, with mainly standard-cell design techniques and portable design flows.

Experimental Results. In direct comparison to a fabricated synchronous chip (Balkan et al., *Hot Interconnects-07*) in the same 90nm technology, our new asynchronous NoC exhibited: *82-91% lower energy/packet, 64-84% less area, and up to 1.7x better system latency than an 800 MHz synchronous network* (for up to 73% of maximum synchronous traffic injection rate), but with some performance degradation at very high traffic rates. Initial simulations of our GALS NoC, in direct comparison with a synchronous architecture, on realistic parallel kernels (array summation, matrix multiplication, breadth-first search, array increment) show promising results. More recently, in developments at Columbia only, with my PhD student Weiwei Jiang, we have demonstrated novel acceleration techniques using early switch pre-allocation*, providing 13-39% further system latency reduction* with almost no additional overhead (ASPDAC-14).

(ii) Targeting General Embedded and Multicore Systems (joint with University of Ferrara, Prof. Bertozzi's group)
This work (2011-present) is in collaboration with a leading NoC group at the University of Ferrara, Italy, with additional work performed only with my Columbia students. The focus is on a more widely used, and more complex, target topology, 2D Mesh (and topologies with higher radix), with the goal of extremely low-power medium-performance embedded systems. However, we also target high-performance multicore systems. The designs are compared directly to a leading low-overhead synchronous NoC, "xpipeLite," using identical technology. There are three key components:

(i) **a basic switch design for 5-ported routers**, with extremely low area and power *(DATE-13)*,

(ii) **a semi-automated CAD tool synthesis flow**, using synchronous tools (Synopsys DC/IC Compiler) *(DATE-13)*, and

(iii) **extensions to support virtual channels (VC's)** *(VLSI-SoC-14)*.

Experimental Results. A post-layout evaluation of the new switch design, in comparison with the synchronous xpipesLite implementation, demonstrated: a *reduction in overall power of 85%/73% (vs. synchronous without/with clock gating), a 71% reduction in switch area, and a 44% reduction in average energy/flit, while maintaining nearly comparable throughput (903 ps/cycle) in a 45nm low power technology.* Work remains to be done on improving and better controlling the tool flow, and on reducing overheads in link-level pipelining. For VC's, we demonstrated that a replicated switch with distinct VC control on links is the best solution. The above designs are almost entirely standard-cell based, making them practical for commercial application. Our solution provides a unique direct comparison with a state-of-the-art synchronous design (xpipesLite), and demonstrates significant overall cost benefits – including highlighting that high-performance asynchronous designs can have significantly lower area than synchronous designs, and can provide much lower average power even than synchronous clock-gated designs.

Selected Publications and Patents:

G. Miorandi, A. Ghiribaldi, S. Nowick and D. Bertozzi, "Crossbar Replication vs. Sharing for Virtual Channel Flow Control in Asynchronous NoCs: a Comparative Study." In *Proceedings of the 22nd IFIP/IEEE International Conference on Very Large Scale Integration and System-on-Chip (VLSI-SoC-14),* Playa del Carmen, Mexico (October 2014).

A. Ghiribaldi, D. Bertozzi and S.M. Nowick, "A Transition-Signaling Bundled Data NoC Switch Architecture for Cost-Effective GALS Multicore Systems." In *Proceedings of the ACM/IEEE Design, Automation and Test in Europe Conference (DATE-13),* Grenoble, France (March 2013) ***(Best Paper Finalist)***.

G. Gill, S.S. Attarde, G. Lacourba and S.M. Nowick, "A Low-Latency Adaptive Asynchronous Interconnection Network Using Bi-Modal Router Nodes." In *Proceedings of the ACM/Symposium on Networks-on-Chip (NOCS)*, Pittsburgh, PA (May 2011).

M.N. Horak, S.M. Nowick, M. Carlberg and U. Vishkin, "A Low-Overhead Asynchronous Interconnection Network for GALS Chip Multiprocessors," *IEEE Transactions on CAD*, vol. 30:4, pp. 494-507 (April 2011) ***(selected for special section on networks-on-chip).***

M.N. Horak, S.M. Nowick, M. Carlberg and U. Vishkin, "A Low-Overhead Asynchronous Interconnection Network for GALS Chip Multiprocessors." In *Proceedings of the ACM/Symposium on Networks-on-Chip (NOCS)"*, Grenoble, France (May 2010).

S.M. Nowick, M.N. Horak and M. Carlberg, "Asynchronous Digital Circuits Including Arbitration and Routing Primitives for Asynchronous and Mixed-Timing Networks," ***U.S. Patent #8,362,802*** (January 29, 2013).

2 CONTINUOUS-TIME DIGITAL SIGNAL PROCESSORS (CT-DSP's)

The goal of this research is targeted to a promising new class of DSP's, which uses variable-rate, i.e. amplitude-based, sampling, called continuous-time DSP's. It is pursued under NSF funding, in collaboration with Prof. Yannis Tsividis (Columbia EE).

The work grows out of early theory, developed and published by Prof. Tsividis, on how DSP's can avoid classic quantization at fixed time intervals (x-axis), and instead adapt their quantized sample rate to conform to the rate of change of the input signal (y-axis). The potential benefits are significant; these include: complete elimination of all aliasing (unlike classic quantization), substantial reduction in overall power (periods of quiescence or low activity inherently result in no or few samples), and the capability of designing high resolution DSP's with lower cost and sample bit-width than classical DSP's.

While a number of prior continuous-time ADC's/DAC's have been developed, ***no flexible and general-purpose CT-DSP has previously been designed.*** The design problems are challenging, since the DSP core must (i) preserve time spacing of input events on its outputs, and (ii) must observe and react in continuous (i.e. not discrete) time to sample arrival. As a result, the CT-DSP's involve an unusual amalgam of *(i) real-time delay-lines and computation*, which translates input time intervals to the outputs precisely, and *(ii) asynchronous digital components and control*, which can handle irregular and unpredictable signal arrival rates, and correctly initiate operation. Only a small handful of prior CT-DSP cores have been implemented, but these are either very early prototypes (e.g. 1-bit sample width, limited sample format support, with timing race conditions not resolved) or specialized for very high throughput (with short delay lines and limited functionality).

In the ESSCIRC-13 and JSSC-14 papers below, ***we demonstrate the first fully-functional general-purpose CT-DSP chip.*** The design, including asynchronous logic and a calibrated delay line, is shown capable of handling a wide variety of input formats (sync PCM, sync/async PWM, sync/async sigma-delta), arbitrary bit widths, and time-varying input rates, with no changes in the DSP design. This capability is not possible in any synchronous DSP without reprogramming. In addition, for certain inputs, it has a signal-to-error ratio which exceeds that of clocked systems. In addition, the frequency response remains intact for any type and rate of its input. The chip also is the first CT-DSP to include on-chip tuning.

Finally, in our ICCD-12 paper, we propose techniques to reduce power in the calibrated delay line. A novel "adaptive granularity" approach is used, with asynchronous circuitry monitoring the current input sample rate, and changing the granularity (i.e. number of pipeline stages) of the delay line accordingly, to reduce overall power (i.e. pipeline depth) for less dense traffic.

Overall, the CT-DSP chip not only makes an important advance in the DSP field; it also highlights the novel and unique benefits that asynchronous design can provide, to enable such an architecture.

Selected Publications and Patents:

Y. Tsividis, M. Kurchuk, S.M. Nowick, B. Schell and C. Vezyrtzis, "Event-Based Data Acquisition and Digital Signal Processing in Continuous Time." Chapter of *Event-Based Control and Signal Processing* (M. Miskowicz, ed.), CRC/Taylor & Francis (2015, *to appear*).

C. Vezyrtzis, W. Jiang, S.M. Nowick and Y. Tsividis, "A Flexible, Event-Driven Digital Filter with Frequency Response Independent of Input Sample Rate," *IEEE Journal of Solid State Circuits* (October 2014, *to appear*).

C. Vezyrtzis, S.M. Nowick and Y. Tsividis, "A Flexible, Clockless Digital Filter." In *Proceedings of the European Solid State Circuits Conference (ESSCIRC-13),* Bucharest, Romania (September 2013).

C. Vezyrtzis, Y. Tsividis and S.M. Nowick, "Designing Pipelined Delay Lines with Dynamically-Adaptive Granularity for Low-Energy Applications." ***Best Paper Award, Logic and Circuit Design Track.*** In *Proceedings of the IEEE International Conference on Computer Design (ICCD-12),* Montreal, Canada (October 2012).

3 ULTRA-LOW ENERGY AND RELIABLE DIGITAL DESIGN

The goal of this research is to develop extremely low-energy digital circuits, using sub-threshold or near-threshold voltage levels, which at the same time provide good performance and high reliability. This work is currently pursued in collaboration with Prof. Mingoo Seok (Columbia EE).

The challenge of low-voltage low-energy digital design is to handle the extremes of device variability and vulnerability to external noise which come with reduced Vdd. Both issues have been heavily studied in the synchronous domain, with much instrumentation and calibration required to mitigate these issues, or else requiring severe margining which limits performance.

Asynchronous design is highly promising for this application, since several design styles have great robustness to timing variability (i.e. using few timing assumptions, and delay-insensitive data encoding), and all eliminate the need for synchronization with a fixed-rate global clock. In this project, in our ISPLED-13 paper, we identify two leading high-performance and highly-robust asynchronous pipeline styles, both using dynamic logic: (i) PS0 (Dean/Horowitz, Stanford), and (ii) PCHB (Lines, Caltech). The former was used in commercial HaL processors in the late 1990's, and the latter is currently used at Intel's Switch & Router Division (formerly Fulcrum Microsystems) for commercial high-performance Ethernet switch chips. We identify a key bottleneck for their use in subthreshold design: "keeper" fighting issues at low voltages**.** We then propose a novel solution: adding lightweight monitoring and adaptive control, to avoid keeper fighting at each pipeline stage before new data is evaluated**.** ***To the best of our knowledge, this is the first approach to demonstrate that asynchronous dynamic pipelines can operate safely at Vdd of 0.3V.*** Our Async-13 paper demonstrates mitigation techniques for correct operation of a leading static asynchronous pipeline, Mousetrap.

Taken together, this initial work is advancing the ability to use robust and high-performance asynchronous pipelines, for extreme low-energy applications.

Selected Publications and Patents:

Y. Chen, M. Seok and S.M. Nowick, "Robust and Energy-Efficient Asynchronous Dynamic Pipelines for Ultra-Low-Voltage Operation Using Adaptive Keeper Control." In *Proceedings of the IEEE International Symposium on Low Power Electronics and Design (ISLPED-13),* Beijing, China (September 2013).

J. Liu, S.M. Nowick and M. Seok, "Soft MOUSETRAP: a Bundled-Data Asynchronous Pipeline Scheme Tolerant to Random Variations at Ultra-Low Supply Voltages." In *Proceedings of the IEEE International Symposium on Asynchronous Circuits and Systems (Async-13),* Santa Monica, CA (May 2013).

4 HIGH-SPEED DIGITAL PIPELINES

The goal of this research is to develop a set of practical asynchronous pipeline circuit structures to support the design of high-performance systems. In synchronous systems, pipelining is the foundation of most high-performance design. Therefore, the development of effective asynchronous pipeline structures is critical to support wide use of clockless design.

We have developed three new asynchronous pipeline structures: ***(i) MOUSETRAP pipelines, (ii) lookahead pipelines,*** and ***(iii) high-capacity pipelines.*** Approach (i) uses static logic implementations, and approaches (ii) and (iii) use dynamic logic implementations. All of the approaches have some common goals: high performance implementations, using simple localized one-sided timing constraints that are easy to satisfy. The static MOUSETRAP pipelines also can be built using existing standard cell libraries. All of these pipelines have been demonstrated to exhibit multi-GigaHertz performance, with low power and modest area overheads.

This work has been influential. *MOUSETRAP pipelines* were adopted by the DARPA CLASS project, led by Boeing Corporation (2005-2007), for use in an experimental asynchronous chip. They were also incorporated by Philips Semiconductors (through their Handshake Solutions startup) into an experimental version of their asynchronous CAD tool flow. A variant of Mousetrap pipelines was developed by NXP/Philips Semiconductors for use in their experimental Aetherial network-on-chip in the early 2000's. Our first paper on *lookahead pipelines* received a Best Paper award at the IEEE Async-00 Symposium. Our *high-capacity pipelines* were adopted by IBM T.J.Watson Research (2001-2002) for use in a fabricated experimental FIR filter chip for disk drive reads (see Technology Transfer below). Four patents have been issued for these pipeline styles.

Selected Publications and Patents:

S.M. Nowick and M. Singh, "High-Performance Asynchronous Pipelines: an Overview." To appear in *IEEE Design & Test of Computers (Sept./Oct. 2011)*

M. Singh and S.M. Nowick, "MOUSETRAP: High-Speed Transition-Signaling Asynchronous Pipelines," *IEEE Transactions on VLSI Systems*, vol. 15:6 (June 2007).

M. Singh and S.M. Nowick, "The Design of High-Performance Dynamic Asynchronous Pipelines: Lookahead Style," *IEEE Transactions on VLSI Systems*, vol. 15:11 (November 2007).

M. Singh and S.M. Nowick, "The Design of High-Performance Dynamic Asynchronous Pipelines: High-Capacity Style," *IEEE Transactions on VLSI Systems*, vol. 15:11 (November 2007).

M. Singh and S.M. Nowick, "High-Throughput Asynchronous Pipelines for Fine-Grain Dynamic Datapaths." ***Best Paper Award.*** In *Proceedings of the IEEE International Symposium on Advanced Research in Asynchronous Circuits and Systems (Async)*, Eilat, Israel (April 2000).

M. Singh and S.M. Nowick, "Circuits and Methods for High-Capacity Asynchronous Pipeline Processing," ***U.S. Patent #7,053,665*** (May 31, 2006).

M. Singh and S.M. Nowick, "Asynchronous Pipeline with Latch Controllers," ***U.S. Patent #6,958,627*** (October 25, 2005).

M. Singh and S.M. Nowick, "Circuits and Methods for High-Capacity Asynchronous Pipeline," ***U.S. Patent #6,867,620*** (March15, 2005).

M. Singh and S.M. Nowick, "High-Throughput Asynchronous Dynamic Pipelines," ***U.S. Patent #6,590,424*** (July 8, 2003).

5 MIXED-TIMING INTERFACES

This aim of this research is to develop a practical and flexible set of interface circuits for mixed-timing digital systems. Heteregeneous systems, which combine asynchronous and/or multiple clock domains, are becoming increasingly common. Support for reliable and efficient interfacing between these domains is a critical capability.

The work defines a complete family of robust circuits that can connect any combination of digital interfaces in different timing domains: clocked-clocked (with different clock rates), clocked-async, and async-clocked. Most future digital systems are expected to have mixed timing domains, yet few practical solutions have been previously proposed to adequately support their interfaces. Our proposed circuits combine several benefits: (i) they can be built using standard gates (no custom circuits required); (ii) they can robustly handle arbitrary timing discrepancies between interfaces (no restriction on the relationship between the different clock rates); (iii) they have high throughput (no synchronization overhead during steady-state operation); and (iv) they are easily scalable (i.e. built using a token ring architecture, with replicated cells).

This work has been influential, with a number of leading architects and network-on-chip researchers (CMU, Intel Barcelona, Cornell, Princeton, Purdue, CEA-LETI, etc.) building on, or using, this approach in their published work.

Selected Publications and Patents:

T. Chelcea and S.M. Nowick, "Robust Interfaces for Mixed-Timing Systems,' *IEEE Transactions on VLSI Systems*, vol. 12:8, pp. 857-873 (August 2004).

T. Chelcea and S.M. Nowick, "Robust Interfaces for Mixed-Timing Systems with Application to Latency-Insensitive Protocols." In *Proceedings of the ACM/IEEE Design Automation Conference (DAC),* Las Vegas, NV (June 2001).

T.Chelcea and S.M.Nowick, "Low-Latency FIFO's for Mixed-Clock Systems." In *Proceedings of the IEEE Computer Society Annual Workshop on VLSI (WVLSI-00),* Orlando, FL (April 2000).

T. Chelcea and S.M. Nowick, "Low-Latency Asynchronous FIFO's using Token Rings." In *Proceedings of the IEEE International Symposium on Advanced Research in Asynchronous Circuits and Systems (Async)*, Eilat, Israel (April 2000).

T. Chelcea and S.M. Nowick, "Low Latency FIFO Circuit for Mixed Clock Systems," ***U.S. Patent #7,197,582*** (March 27, 2007).

T. Chelcea and S.M. Nowick, "Low Latency FIFO Circuits for Mixed Asynchronous and Synchronous Systems," ***U.S. Patent #6,850,092*** (February 1, 2005).

6 ROBUST ENCODING FOR GLOBAL ASYNCHRONOUS COMMUNICATION

This aim of this research is to develop practical codes, and hardware support, for robust asynchronous communication. These codes can be used for communication either in fully-asynchronous systems, or in GALS systems which allow use of synchronous cores, memories and function units integrated through a flexible asynchronous network. This work is centered on ***delay-insensitive (DI) codes*** (also known as "unordered codes" in the synchronous community). The benefit of these codes is that they tolerate arbitrary skew in the arrival of individual bits in a transmission, and hence support *timing-robust* communication.

The research includes the design of several new practical DI codes: **(i) Zero-Sum**, which also provides error correction (i.e. *an error-correcting unordered [ECU] code*); **(ii) DI Bus-Invert**, which provides low power through use of selective bit inversion; and **(iii) Level-Encoded Transition Signalling (LETS)**, which provides both high-throughput (using a two-phase protocol) and low power. The work on Zero-Sum codes also includes heuristic extensions to provide good coverage of 2-bit correction while guaranteeing 100% coverage of 1-bit correction. In addition, supporting hardware blocks for protocol conversion, encode/decode and completion detection have been developed. Protocol conversion allows efficient encodings for global channels to interface to alternative efficient encodings for local computation nodes. Completion detectors are critical components of asynchronous systems: they determine when a valid DI codeword has been received.

LETS codes have recently been successfully used in the Stanford University "Neurogrid" project (led by Prof. Kwabena Boahen), to build a leading large-scale neuromorphic system, where low-power, high-throughput and robust inter-neuron communication is facilitated using asynchronous DI communication (see Sec. V.B. of *Proceedings of the IEEE*, vol. 102:5, pp. 699-716, April 2014).

Selected Publications and Patents:

M.Y. Agyekum and S.M. Nowick, "Error-Correcting Unordered Codes and Hardware Support for Robust Global Communication," *IEEE Transactions on Computer-Aided Design,* vol. 31:1, pp. 75-88 (January 2012).

M.Y. Agyekum and S.M. Nowick, "A Delay-Insensitive Bus-Invert Code and Hardware Support for Robust Global Communication." In *Proceedings of the ACM/IEEE Design, Automation and Test in Europe Conference (DATE-11),* Grenoble, France (May 2011).

M. Cannizzaro, W. Jiang and S.M. Nowick, "Practical Completion Detection for 2-of-N Delay-Insensitive Codes." In *Proceedings of the IEEE International Conference on Computer Design (ICCD-10),* Amsterdam, The Netherlands (October 2010).

M.Y. Agyekum and S.M. Nowick, "An Error-Correcting Unordered Code and Hardware Support for Robust Global Communication." In *Proceedings of the ACM/IEEE Design, Automation and Test in Europe Conference (DATE),* Dresden, Germany (March 2010).

A. Mitra, W.F. McLaughlin and S.M. Nowick, "Efficient Asynchronous Protocol Converters for Two-Phase Delay-Insensitive Global Communication," *IEEE Transactions on VLSI Systems*, vol. 18:7, pp. 1043-1056 (July 2009).

P.B. McGee, M.Y. Agyekum, M.A. Mohamed and S.M. Nowick, "A Level-Encoded Transition Signaling Protocol for High-Throughput Asynchronous Global Communication." In *Proceedings of the IEEE International Symposium on Advanced Research in Asynchronous Circuits and Systems (Async)*, Newcastle-upon-Tyne, UK (April 2008).

7 COMPUTER-AIDED DESIGN TOOLS

A key goal of this work is to develop practical CAD tools for asynchronous design. CAD tools are critical to the widespread adoption of asynchronous circuits.

My PhD thesis research, and accompanying publications, presented some of the earliest solutions for asynchronous CAD design and optimization techniques and tool development, targeting asynchronous controllers, called ***"locally-clocked" burst-mode machines***. These asynchronous state machines were the first to guarantee correct hazard-free operation when simultaneously targeting: implementations with realistic gate-level mapping, supporting a fairly general specification style, yet still providing very low latency and area overhead. The thesis also introduced the first general solution to the 2-level minimization problem for hazard-free logic.

More recently, my focus has been on four key areas: (i) individual controllers, (ii) timing-robust circuit styles (i.e. threshold circuits), (iii) performance analysis and timing verification of concurrent systems, and (iv) automed synthesis and optimization of large-scale asynchronous systems.

I have recently released new versions of the first three of these tools under a comprehensive asynchronous design framework, called "**CaSCADE"** (Columbia University and USC Asynchronous Design Environment). Each tool is available for free download, and includes extensive tutorial support, setup documentation and benchmark examples. For download access, or just to view tutorial slides, see: http://www.cs.columbia.edu/~nowick/asynctools.

(i) MINIMALIST: a CAD Package for Synthesis & Optimization of Asynchronous Controllers
With several of my students, I have developed and released a large software CAD package for the synthesis of asynchronous burst-mode controllers, called MINIMALIST.

"Burst-mode" asynchronous controllers have a number of attractive features: simple timing constraints, very low latency (i.e., input-to-output paths) and power consumption, and targeting of existing (synchronous) standard-cell gate libraries for ease of implementation. They have been successfully applied to a variety of industrial and large-scale designs, e.g. cache and SCSI controllers, and controllers for an experimental low-power infrared communication chip from HP Laboratories ("Stetson" project). Recently, burst-mode controllers and the MINIMALIST package have been used at NASA Goddard Space Flight Center to design experimental space measurement chips (see "Technology Transfer" below).

The Minimalist CAD package provides a number of practical benefits for designers. It includes: ***(a) a sophisticated set of optimization algorithms*** (including both exact and heuristic) for several synthesis steps: two-level and multi-level hazard-free logic minimization, and optimal critical race-free state assignment (under a so-called 'input encoding' model) – where we have proposed the first complete solutions to these problems [ICCAD-92, TCAD-95, ICCAD-95]; ***(b) an extensive set of designer scripts,*** allowing users to target different cost functions (speed vs. area), typically providing dozens of alternative implementations synthesized under different user metrics, thus supporting design-space exploration; ***(c) extensive practical designer support,*** including a Verilog backend, graphical interfaces, a command-line shell (for customized synthesis runs), and automatic insertion of initialization circuitry; and ***(d) a top-to-bottom verifier,*** to validate the final implementation against the initial specification.

The tool has been highly visible: it has been downloaded to over 100 sites in over 18 countries. It has also been used in a joint project between the nominee and NASA Goddard Space Flight Center (see below). Portions of the tool (hazard-free two-level logic minimization) have been used in experimental projects at HP Laboratories (Stetson project) and also incorporated into other asynchronous CAD tools (3D tool from Kenneth Yun).

Selected Publications and CAD Tools:

R.M. Fuhrer and S.M. Nowick, *Sequential Optimization of Asynchronous and Synchronous Finite-State Machines: Algorithms and Tools.* Kluwer Academic Publishers, Boston, MA (2001).

L. Lavagno and S.M. Nowick, "Asynchronous Control Circuits". Chapter 10 of *Logic Synthesis and Verification* (pp. 255-284) ***(invited contribution).*** S. Hassoun and T. Sasao, editors. Kluwer Academic Publishers, Boston, MA (2002).

M. Theobald and S.M. Nowick, "Fast Heuristic and Exact Algorithms for Two-Level Hazard-Free Logic Minimization," *IEEE Transactions on Computer-Aided Design,* vol. 17:11, pp. 1130-1147 (November 1998).

R.M. Fuhrer and S.M. Nowick, "Symbolic Hazard-Free Minimization and Encoding of Asynchronous Finite State Machines." In *Proceedings of the ACM/IEEE International Conference on Computer-Aided Design (ICCAD),* San Jose, CA (November 1995).

S.M. Nowick and D.L. Dill, "Exact Two-Level Minimization of Hazard-Free Logic with Multiple-Input Changes," *IEEE Transactions on Computer-Aided Design,* vol. 14:8, pp. 986-997 (August 1995).

See also tutorial slides and CAD tool download information available at http://www.cs.columbia.edu/~nowick/asynctools.

(ii) Synthesis and Optimization of Asynchronous Threshold Circuits

Another research area has been to develop CAD tools for a class of extremely timing-robust asynchronous circuits: ***dual-rail threshold circuits***. These circuits are especially important for future-generation digital systems, because they gracefully tolerate wide variations and unpredictability due to process, temperature and voltage variability. However, there has been only limited previous work on developing systematic optimization techniques for these circuits. The goal is to support automated design of substantial digital systems (datapath + control) using this design style.

A complete tool package, called **ATN_OPT,** is available on the CaSCADE web site, with a tutorial; see: http://www.cs.columbia.edu/~nowick/asynctools.

In particular, my students and I have developed novel optimization algorithms and CAD tools for two synthesis steps: ***(i) technology mapping (atn_map)***, and ***(ii) multi-level optimization (atn_relax)***. The work demonstrates the feasibility of powerful CAD optimizers for robust asynchronous circuits, which can obtain *over 50% performance and area improvement* without any loss of their timing-robustness properties.

The technology mapper, ***atn_map***, is the *first systematic and general approach* for technology mapping of asynchronous threshold circuits. It is based loosely on synchronous techniques (such as those incorporated into Synopsys' Design Compiler), but with novel modifications to several steps to ensure that no hazards or timing constraints are introduced. A basic method has been developed to handle individual cost functions (area, delay). A recent extension has been developed to handle an important combined cost function: *area optimization under hard delay constraints*. The latter is the

first asynchronous logic synthesis approach to systematically ensure that hard timing requirements are met.

The multi-level optimizer, ***atn_opt***, is based on the notion of "local relaxation". The strategy is to implement selected nodes in a given netlist using "eager evaluation", i.e. allowing these nodes to produce outputs *early (*before all inputs have arrived). Other nodes in the netlist are implemented in a conservative "input-complete" manner: only producing outputs after all inputs have arrived. If the relaxed nodes are carefully selected, this hybrid 'relaxation' approach *still* ensures a fully timing-robust overall circuit.

These tools have been applied to substantial systems with thousands of gates and hundreds of inputs and outputs: a complete DES encryption circuit, a GCD circuit, and the largest MCNC combinational benchmarks. For technology mapping, *average delay improvements of 31.6% and area improvements of 9.5% were obtained.* For multi-level optimization, *average delay improvements of 16.1% and area improvements of 34.9% were obtained.* Using our recent technology mapping approach targeting delay-area tradeoffs, as a post-processing step, *further area reduction by 10.7% on average can be recovered*, with no degradation of delay. The resulting circuits still remain hazard-free and timing-robust. This approach has also extended this work to handle simple *combinational* dual-rail circuits, i.e. dual-rail circuits that do no have hysteresis. This circuit family is widely used, and though less timing-robust than sequential threshold circuits, provides higher-performance circuits. The optimization methods can apply directly to this larger class.

Publications and CAD Tools:

C. Jeong and S.M. Nowick, "Technology Mapping and Cell Merger for Asynchronous Threshold Networks," *IEEE Transactions on Computer-Aided Design*, vol. 27:4, pp. 659-672 (April 2008).

C. Jeong and S.M. Nowick, "Optimization for Timing-Robust Asynchronous Circuits based on Eager Evaluation." In *Proceedings of the IEEE International Symposium on Asynchronous Circuits and Systems (Async),* Newcastle-upon-Tyne, UK (April 2008).

C. Jeong and S.M. Nowick, "Optimization of Robust Asynchronous Circuits by Local Input Completeness Relaxation." In *Proceedings of the Asia-South Pacific Design Automation Conference (ASPDAC)*, Yokohama, Japan (January 2007).

C. Jeong and S.M. Nowick, "Optimal Technology Mapping and Cell Merger for Asynchronous Threshold Networks." In *Proceedings of the IEEE International Symposium on Asynchronous Circuits and Systems (Async)*, Grenoble, France (March 2006).

C. Jeong and S.M. Nowick, "Methods, Media and Means for Forming Asynchronous Logic Networks," ***U.S. Patent #7,729,892*** (June 1, 2010).

See tutorial slides and CAD tool download information available at
http://www.cs.columbia.edu/~nowick/asynctools.

(iii) Performance Analysis and Timing Verification of Concurrent Systems

A key component of an effective design methodology is to have efficient performance analysis and timing verification tools, both to evaluate a system's operation and also as a basis for driving further optimization. Performance analysis and timing verification are especially challenging for asynchronous systems: since there is no global clock, and system-level operation is often highly decoupled and concurrent, it is difficult to adapt existing synchronous analytical or simulation approaches.

In this work, we have developed two novel approaches to performance analysis: ***(a) using stochastic delay models*****,** and ***(b) using min/max delay models***. A complete tool package, called **DES (Discrete Event System) Analyzer,** is available on the CaSCADE web site, with a tutorial; see: http://www.cs.columbia.edu/~nowick/asynctools.

Approach (a), called ***des_perf***, focuses on modeling delay distributions for each component in a concurrent system, and then using Markovian techniques to find key metrics: average-case throughput, latency, relative input arrival order and component utilization. Critical and slack paths can also be identified. These metrics can then be used either to rate the system's performance, or as a basis for performance-driven optimization of the system.

Approach (b), called ***des_tse***, does not consider a delay distribution, but rather extremes of behavior: min and max delays of each component. It then evaluates the 'time-separation' between two successive events in the system. This approach can then be used to rapidly determine relative orderings of input arrivals and extremes of overall system behavior. It is therefore a basis for formal verification of feasible behaviors.

Selected Publications:

P.B. McGee and S.M. Nowick, "An Efficient Algorithm for Time Separation of Events in Concurrent Systems." In *Proceedings of the International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA (November 2007).

P.B. McGee and S.M. Nowick, "Efficient Performance Analysis of Asynchronous Systems Based on Periodicity." In *Proceedings of the IEEE/ACM International Conference on Hardware/Software Codesign and System Synthesis (CODES),* Jersey City, NJ (September 2005).

See also tutorial slides and CAD tool download information available at
http://www.cs.columbia.edu/~nowick/asynctools.

(iv) Synthesis and Optimization of Large-Scale Asynchronous Systems

A final area has been to develop CAD tools for the synthesis and optimization of entire asynchronous systems.

Two approaches have been developed. With my PhD student Tiberiu Chelcea and colleagues at the University of Manchester, we have developed a "back-end optimizer" suitable for improving the existing asynchronous commercial CAD flow at Philips Semiconductors. The Philips approach uses simple syntax-directed compilation, from a high-level concurrent specification language to asynchronous circuits. In this approach, each construct in the specification is replaced directly by a corresponding small "process" (i.e. a concurrent component) in hardware. Our proposed approach is to build on this unoptimized compiler flow, and insert a powerful back-end optimizer, which performs peephole and resynthesis transformations on the netlist of concurrent hardware components, and thereby further restructures and improves the circuit. Initial results of our tool indicate up to 50% performance improvements.

An alternative approach, developed with my PhD student Michael Theobald, is closer to classic high-level synthesis from the synchronous world, but introduces concurrency-enhancing system-level transformations, in a strict architecture where each function unit has a dedicated asynchronous controller.

Selected Publications:

T. Chelcea and S.M. Nowick, "Resynthesis and Peephole Transformations for the Optimization of Large-Scale Asynchronous Systems." In *Proceedings of the IEEE/ACM Design Automation Conference (DAC)*, New Orleans, LA (June 2002).

M. Theobald and S.M. Nowick, "Transformations for the Synthesis and Optimization of Asynchronous Distributed Control." In *Proceedings of the IEEE/ACM Design Automation Conference (DAC)*, Las Vegas, NV (June 2001).

8 TECHNOLOGY TRANSFER

Finally, I have been involved in several efforts to promote asynchronous technology transfer. (For other highlights, see DARPA CLASS project under "Grants" below.)

(i) NASA Goddard Space Flight Center: Space Measurement Applications (2006-2008)

I recently collaborated with NASA Goddard Space Center in the design of series of asynchronous chips for space measurement applications (Spring 2006 to present).

In Spring 2006, I was invited by a senior engineer at NASA/Goddard (now a manager) to collaborate in designing an asynchronous measurement circuit for space applications. The motivation for using asynchronous design are: the complete removal of a high-speed sampling clock, lower power, and design flexibility in implementing a highly-concurrent and fine-grained micro-architecture which can handle varied input sampling rates.

The chip includes a number of my asynchronous burst-mode controllers, and the design made extensive use of my *Minimalist* CAD package. The asynchronous circuit was designed almost entirely by myself and my NASA colleague, with additional support for simulation and physical design by NASA engineers. The design has been included in a fabricated chip; it is expected to achieve desired performance targets but with significantly lower area and power than their previous synchronous designs. This work has received positive interest from NASA scientists, and there is some possibility of its use in future space missions.

(ii) IBM T.J. Watson: FIR Filter Chip for Disk Drive Reads (2001-2002)

My former PhD student (Montek Singh) and I transferred one of our high-speed asynchronous pipeline styles to IBM, for use in an experimental FIR filter chip for disk drive reads.

This work presents an important advance in the design of high-speed asynchronous circuits: it demonstrates that an *industry-standard FIR filter* could be designed by mixing asynchronous pipelines and synchronous interfaces, achieving *higher-throughput* and *significantly-lower latency* than the *best* comparable commercial synchronous design.

The chip was designed by my PhD student, Montek Singh, and IBM engineers, including Dr. Jose Tierno and others. The entire core of the filter was asynchronous -- designed using our *high-capacity dynamic pipelines* --, while the external interfaces were synchronous. Hence, the chip appears to the environment as a fully synchronous system. The chip was designed to IBM commercial specifications.

The fabricated chip was evaluated in April-June 2001; it was fully functional. The synchronous interfaces operated at 1.3 Gigasample/sec (in 0.18 micron technology), but the internal asynchronous

pipelines (using our asynchronous dynamic pipelines) were shown to support a much higher rate: 1.8 Gigasample/sec. The chip *met and exceeded the synchronous performance* and power requirements for the next-generation process at IBM; it also obtained 15% higher throughput than current-generation synchronous chips. Most significantly, the new hybrid design had the advantage over synchronous ones of *dynamically-variable latency*, depending on input sampling rate. Hence*, the average latency was significantly lower than for the best comparable commercial synchronous IBM chips.*

Selected Publications:

M. Singh, J.A. Tierno, A. Rylyakov, S. Rylov, and S.M. Nowick, "An Adaptively-Pipelined Mixed Synchronous-Asynchronous Digital FIR Filter Chip Operating at 1.3 GigaHertz," *IEEE Transactions on VLSI Systems*, vol. 18:7, pp. 1043-1056 (July 2010).

J. Tierno, A. Rylyakov, S. Rylov, M. Singh, P. Aspadu, S.M. Nowick, M. Immediato, and S. Gowda, "A 1.3 GSample/s 10-tap Full-rate Variable-Latency Self-Timed FIR with Clocked Interfaces," in *Proceedings of the International Solid State Circuits Conference (ISSCC),* Monterey, CA (February 2002).

M. Singh, J.A. Tierno, A. Rylyakov, S. Rylov, and S.M. Nowick, "An Adaptively-Pipelined Mixed Synchronous-Asynchronous Digital FIR Filter Chip Operating at 1.3 GigaHertz" (***Best Paper Finalist)***. in *Proceedings of the IEEE International Symposium on Advanced Research in Asynchronous Circuits and Systems (Async)*, Manchester, UK (April 2002).

9 GRANT HIGHLIGHTS (2000-2014)

Below are some highlights of recent grants (see CV for additional grants).

(i) NSF Award (medium-scale)

Title: ``***SHF:Medium: Power-Adaptive, Event-Driven Data Conversion and Signal Processing Using Asynchronous Digital Techniques"***

Time Period: 7/1/10-1/31/15
Total Grant Amount: $1,062,605
My Grant Portion: $500,000 *(approx.)*

Principal Investigator: Prof. Yannis Tsividis, EE Dept., Columbia U. (joint with co-PI S. Nowick)

This medium-scale NSF award targets ultra low-power microelectronic systems, through the development of a new approach to digital signal processing and conversion, called *continuous-time digital signal processing (CT-DSP)*. This work is aimed at applications that require continuous monitoring and processing of information, which may arrive infrequently, or at irregular or unpredictable intervals. Application areas include environmental sensors, and implantable or ingestible biomedical devices. Traditional synchronous processing is a poor match for these applications, because the regular sampling and clock control result in excessive power and aliasing. The event-based nature of the information calls for a drastic re-thinking of how these signals are monitored and processed. This research aims instead to provide a system controlled not by the clock, but by the actual arrival of each event. Asynchronous (i.e. clockless) digital logic techniques, which are ideally suited for this work, are combined with continuous-time data conversion and digital signal processing. This new "event-based" approach promises significant power and energy reduction, as well as higher-quality sampling, in a fully-programmable chip.

(ii) NSF Award (medium-scale)

Title: ``***CPA-DA-T: Design and Tools for Easy-to-Program Massively Parallel On-Chip Systems: Deriving Scalability Through Asynchrony"***

Time Period: 8/1/08-7/31/14
Total Grant Amount: $921,686
My Grant Portion: $461,000

Principal Investigator: Prof. Steven Nowick (joint with co-PI Prof. Uzi Vishkin, U. of Maryland)

This medium-scale NSF award is to support the design and optimization a high-throughput, flexible and low-power digital interconnection network for future desktop parallel processors. The asynchrony will facilitate lower power, handling of heterogeneous interfaces, and high access rates (with fine-grained pipelining). This work is in collaboration with the parallel processing architecture group at the University of Maryland, which is developing the synchronous shared-memory processor environment

(called XMT) and simulation tools. My group is leading the design of the novel asynchronous interconnection network. The goal is a *globally-asynchronous locally-synchronous (GALS) network*, that can integrate multiple synchronous cores and caches operating at unrelated clock rates. Most recent GALS networks have focused on low- to medium-performance embedded systems, or involve advanced circuit techniques for high-performance systems. In contrast, this proposal is aimed at medium-to-high end multi-processors, but using largely standard-cell design techniques, and portable design flows.

(iii) DARPA "CLASS" Project (MTO)

Principal Investigator/Lead: Boeing Corporation

Time Period: Fall 2005 through March 2007
Total Contract Amount: $14,000,000
My Subcontract Portion: $502,000

This DARPA program is the largest US government research program for asynchronous digital design in the last 30 years. Its goal is to make asynchronous digital design viable for the commercial and military sectors. There were approximately 20 large-scale proposals submitted, and *only 1 contract funded*, headed by Boeing Corporation (PI), with participation of Philips Semiconductors (via its incubated asynchronous startup company, called Handshake Solutions), two asynchronous startups (Theseus Logic, Codetronix) and three key academic groups (Columbia, UNC, U. of Washington).

The two goals of the project are: (i) building a large-scale asynchronous demonstration chip (for Boeing military applications), and compare its performance and cost to an equivalent synchronous chip; and (ii) provide a "legacy asynchronous CAD tool" for future asynchronous designs.

I was brought onto the project at the start of Phase 2, to provide expertise in CAD tool development and optimization techniques. My role was to provide three key components: (i) CAD optimization techniques for robust asynchronous threshold circuits (to improve the unoptimized tool flow provided by Theseus Logic); (ii) CAD optimization techniques for the Philips-based asynchronous tool flow (collaborating with a team from its incubated startup company, Handshake Solutions); and (iii) migration our high-speed MOUSETRAP asynchronous pipelines into the tool flow.

(iv) NSF ITR Award (medium-scale)

Title: ***"A CAD Framework for the Design and Optimization of Large-Scale Asynchronous Digital Systems"***

Time Period: 9/1/00-8/31/07
Total Grant Amount: $1,600,000
My Grant Portion: $ 806,000

Principal Investigator: Prof. Steven Nowick (joint with co-PI Prof. Peter Beerel, USC).

(v) NSF ITR Award (medium-scale)

Title: ***"Asynchronous Digital Signal Processing for the Software Radio"***

Time Period: 9/1/00-8/31/03
Total Grant Amount: $969,227
My Grant Portion: $400,000 *(approx.)*

Principal Investigator: Prof. Ken Shepard, EE Dept., Columbia U. (joint with co-PI S. Nowick)

In grant (iv), we have developed a comprehensive CAD tool framework for the synthesis and optimization of asynchronous systems, targeting different asynchronous circuit styles (from high-speed/less robust to moderate-speed/highly-robust). This work encompasses the entire ***CaSCADE Tool Environment*** free public-domain tool release, available on the web site: www.cs.columbia.edu/~nowick/asynctools. Columbia tools include: (i) *MINIMALIST,* for synthesis and optimization of asynchronous burst-mode controllers; (ii) *ATN_OPT,* for asynchronous threshold networks; and (iii) *DES Analyzer,* for performance analysis and verification of concurrent systems. (See descriptions above.)

In grant (v), we explored the applicability of high-speed asynchronous pipelines to designing voltage-controlled asynchronous micro-architectures for software radio.

Note: The ITR initiative is an outgrowth of President Clinton's "PITAC" advisory committee for national information technology, to fund "long-term risk-taking research" in information technology. In 2000, only 62 medium-scale NSF ITR awards were granted, out of 920 submitted proposals, across all areas of information technology. *I was 1 of only 4 investigators nationally in 2000 to receive 2 medium-scale NSF ITR awards.*