# System Level Design
# Research Needs: System Level Design
# April 2016

The Semiconductor Research Corporation (SRC) Global Research Collaboration (GRC) program member companies are pleased to release this document that describes the research needs in the System Level Design thrust. Incorporated into this document are the needs identified through SRC GRC Executive Technical Advisory Board (ETAB) priorities, the System Level Design (SLD) TAB strategic planning process and other discussions.

In the preparation of this document three themes emerged as key priorities. System optimization, system validation (including power-performance aspects), and system firmware/software (where the focus is on SW/HW co-development).

The system design research needs of the members are described in nine major categories:

- Early Design Space Exploration at System Level
- System Level Architectures for Implementation and Validation
- Design Robustness
- System Power Optimization
- Multi-core SoC Design
- End-to-End Data Aggregation and Processing
- Algorithms for Sensing and Real Time Embedded Control/Communication Applications
- SW/HW Co-development
- Security and Anti-Counterfeiting

Each of these major categories are broken down into several sub-categories which describe the need in more detail. Even so, these are written to be broad in nature to not restrict the investigator's approach. The sub-categories that are in bold text indicate that there is higher common interest among the members but that may not necessarily reflect a higher priority in the final selections.

The needs in the system design space cover a broad range of applications, including high performance processors for data centers, automotive, industrial, mobile computing and communication, healthcare, and efficient energy usage and management systems. Investigators are encouraged to link the results of their work with a potential application to help show the relevance of the proposed work.

This needs document is driving the System Level Design solicitation. It is issued to universities worldwide, may be addressed by an individual investigator or a research team. Our selection process is divided into two stages. The interested party is requested to submit a brief 1-page white paper. The white paper should clearly identify what can be done in two years, as well as what additionally could be done if a third year is requested. (Please specify the third-year goals separately.) Two-year-only white papers are also acceptable. A successfully selected white paper will result in an invitation to submit a full proposal, with recommendation that the proposal will be written for 2 years or 3 years. These proposals will be further down-selected for research contracts. The number and size of the contracts awarded will be determined by the amount of available funds, and by the number of high-quality proposals

Investigators who are funded will be expected to publish at top-tier conferences, including but not limited to ISSCC, VLSI, HPC, ISCA (part of Federated Computing Research Conference), and ESWEEK (CASES, CODESISSS, & EMSOFT).

White Papers for all the categories below will be considered for funding. Note that there is no priority assigned to the order of the major categories or the sub categories. Investigators are limited to participation in two white papers in this System Level Design solicitation. Submissions should highlight major category needs are addressed, such as "S2".

**CONTRIBUTORS**

| | |
|---|---|
| ARM | Jose Joao |
| Freescale | Sourav Roy, John Pigott |
| GLOBALFOUNDRIES | Ali Wehbi |
| IBM | Prabhakar Kudva |
| Intel | Michael Kishinevsky, Arindam Saha |
| Mentor Graphics | Serge Leef |
| TI | Anand Dabak, Steven Bartling |

# 2016 Integrated System Design Needs Categories

| | |
|---|---|
| **S1** | **Early Design Space Exploration at System Level** |
| *The increasing inter-dependencies of primary system design objectives such as power, performance, reliability, security, and programmability are simultaneously complicated by technology trends toward new memory devices (e.g., NVRAM), new packaging options (e.g., die stacking), new interconnect technologies (e.g., photonics), and the inherent complexity from continuously increasing the number of devices per system. Research needs include developing modeling methods, tools, and algorithms, and exploring systems, architectures, and sub-systems to enable effective systems that exploit (or mitigate) technology trends.* | |
| S1.1 | Algorithms, methods, and models for quantifying performance, power, reliability, and cost trade-offs in early design |
| S1.2 | Exploration and estimation for run-time adaptive heterogeneous systems |
| S1.3 | High level modeling, design and synthesis flows to enable early analog system design and tradeoff analysis for analog sub-systems |
| **S1.4** | **Exploration of architecture and systems that incorporate emerging memory types** |
| **S2** | **System Level Architectures for Implementation and Validation** |
| *Demanding applications from data centers to smart devices continue to drive systems with improved metrics as measured with multiple figures of merit. Of high value are research tasks that show novel architectures, communication, and microarchitectures for processor cores, memory subsystems, analog/RF subsystems, and other building blocks that display significant improvement in metrics such as performance, power and energy consumption, robustness, reconfigurability, functionality, and design/implementation costs. Of high value in this category are architectures and methods that address post-silicon validation issues.* | |
| **S2.1** | **System level validations of performance, power and energy consumption, and robustness including both hardware and software** |
| S2.2 | Analog/RF/mixed-signal sub-systems architectures, including ultra-low power applications |
| **S2.3** | **Effective and fast turn-around design flows for SoC/SiP/3D, including high performance IP blocks (e.g. processor cores)** |
| S2.4 | System architectures having access to significantly higher off-chip bandwidth (e.g. optical) |
| S2.5 | Microarchitecture synthesis including customized reconfigurable accelerators |
| S2.6 | System correctness and optimization in aggressive power and energy management schemes |
| **S2.7** | **Novel memory sub systems and architectures enabled by the inclusion of emerging non-volatile memory technologies** |
| S2.8 | High-performance processor microarchitectures |
| S2.9 | Architectural options for SoCs integrating reconfigurable hardware and interconnect at various levels of granularity |
| **S2.10** | **System architectures and methods that address post silicon validation issues including design/implementation cost** |
| **S3** | **Design Robustness** |
| *Reliability continues to be a significant system-design concern. End-to-end computing and stacked nature of large scale systems, e.g., from sensors to cloud data analytics/warehouses increases complexity, and increasing number of potentially weak points for reliability. Continuing primary research foci include estimation/modeling, avoidance, detection, recovery, correction, containment, and repair of soft/transient and hard/permanent failures, plus failures from gradual deterioration and communication between components. A primary focus on end-to-end and cross-layer reliability provides new opportunities for innovation.* | |
| S3.1 | Low-power, low-cost fault-detection, -correction, -containment, and –repair micro/architecture for end-to-end systems |
| S3.2 | System-repair architectures with quantifiable performance degradation, at reasonable power and cost overhead |
| S3.3 | Highly available distributed and cloud architectures at substantially reduced power and cost overhead |
| S3.4 | Evaluation models for fault-tolerant system architectures for end-to-end data integrity |
| S3.5 | Novel architecture, microarchitecture, software, and communication protocols that tolerate failures along end-to-end data flows. |

| | |
|---|---|
| **S4** | **System Power Optimization** |
| *One of the major concerns in integrated system design is power management and energy minimization. Traditional performance-power trade-off techniques are inadequate to address future technology leakage power and computational demands in a reliable manner. Multicore and heterogeneous architectures will be the norm. Future power gains will likely be made by the addition of dedicated accelerators and reconfigurable hardware. Software and hardware cooperation is needed to optimize use of resources in different parts of the system. Novel global optimization techniques will be needed at every design step, top to bottom, to achieve the targeted solution, addressing power for computation, memory and on-die and on-package communication.* | |
| **S4.1** | **Efficient static and dynamic power management and optimization techniques for SoC and platforms from nano-watt to high-end systems** |
| S4.2 | Novel techniques for reducing on-die and off-die data communication power |
| S4.3 | Reducing di/dt through microarchitectural design and software techniques (application, OS, and firmware), and run-time management |
| **S4.4** | **Novel techniques for resource scheduling and workload balancing** |
| **S5** | **Multi-core SoC Design** |
| *Multi-core, many-core, and heterogeneous processor designs provide ways of continuing performance scaling while remaining within the limitations of fixed power, energy, and thermal budgets. New architectures and design methodologies for such processors will be needed to continue providing more performance across a wide range of form factors, performance targets, and power constraints. In addition to novel processor organizations and accelerators, further research is needed in scalable and flexible interconnection fabrics and corresponding protocols to tie together the many different components of future SoCs. At the same time, emerging die-stacking and system-in-package technologies provide new opportunities (mixed process technologies, dense integration, high chip-to-chip bandwidths, IP reuse) to innovate novel processor and interconnect architectures, and research is required to determine the most promising approaches.* | |
| S5.1 | Multi-core or many-core (homogeneous and/or heterogeneous) architectures and design methodologies for micro watt to high power systems |
| S5.2 | On-chip communication fabrics and protocols that provide significant improvement over state of the art |
| S5.3 | SiP/3D chip-to-chip architectures/interfaces and design methodologies |
| S5.4 | Hardware-software solutions to the problem of parallelizing single-threaded algorithmic flows |
| S5.5 | Hardware-Software solutions for more efficient execution of parallel applications |
| **S6** | **End to End Data Aggregation and Processing** |
| *End-to-end data aggregation and processing (from data source to destination, e.g., sensors to gateways to cloud to end-users or data analytics engines), with co-design of hardware, software and middleware. The end-to-end path may include multiple devices, paths and protocols with sensors, embedded devices, networks, compute and storage locations. New architectures and methods for improved performance, power, security and reliability across the span of an end-to-end system create new opportunities for innovation. Both data and compute centric systems, including internet of things (IoT) with different response time requirements, from real time to batch processing, are increasingly important. Novel architecture co-design across devices, interconnection paths, and protocols between the devices as well as computational hierarchies (i.e. end-node vs. fog vs. cloud) need to be considered.* | |
| S6.1 | End-to-end systems: data flow, device/component and path aware co-design of hardware, embedded software, middleware and systems for power, performance and reliability |
| S6.2 | Virtualization, programmability, runtimes and dynamic modification of end-to-end hardware and software systems |
| S6.3 | Modeling and simulation of end-to-end data flows based on application domains for data aggregation and processing (eg. Mobile health, safety) |
| **S6.4** | **Use of novel system architectures enabled by the inclusion of emerging technologies: non-volatile memories, high speed IO and networking, non von-Neumann and cognitive hardware for such aggregation and processing** |
| S6.5 | IT system architectures for big data with decentralized processing approaches in addition to centralized data centers |
| S6.6 | Compute, network, protocol, storage and system design for distributed real-time interactive, mission-critical and deep learning and long-term analytics |

| | |
|---|---|
| **S7** | **Algorithms for Sensing and Real Time Embedded Control / Communication Applications** |
| *This category pertains to signal processing related to data acquisition and control in sensing applications wherein power, cost, and computational throughput per quanta of energy consumption (at usable levels of performance) are key metrics. Signal processing is loosely defined in this context and could cover a wide range of actions such as filtering, algorithmic compensation for hardware limitations, trend analysis, etc., as well as meta-level analysis and control, in which data is aggregated from lower level sensing elements to present a coherent view that provides enhanced functionality. Applications have an equally wide span and include but are not limited to industrial sensing and control (including motor control), commercial and residential sensing/control and health/medical sensing, automobile, autonomous vehicles, and smart grid applications.*<br><br>*Industrial sensing includes such areas as 20mA loop sensing and control applications, gas or liquid flow sensing, and process monitoring. The smart grid arena includes electric power grid monitoring/control, metering, infrastructure monitoring (traffic lights, building, bridges, street lighting, gas or water flow metering, wind and solar generators, etc.). Health/fitness/medical monitoring include wireless sensing/analysis for ambulatory medical monitoring, fitness parameters such as heart rate, blood pressure, activity level, ECG. Residential/commercial applications would include sensing occupancy, temperature, relative humidity, lighting level, security related sensing (glass breakage, motion, etc.) and control of HVAC, lighting, etc. Automotive applications include sensor fusion, low-end vision, and wireless in-vehicle networks.* | |
| S7.1 | Sensors and algorithms/IP cores enabling combining purposely different forms of sensor data (e.g. stereoscopic imaging + time-of-flight phase info + imaging [mm Wave or THz]) |
| S7.2 | Multi-core algorithms, coding, and software tools for embedded ultra-low power sensing and control applications |
| S7.3 | Algorithms and architectures for energy efficient embedded vision (including 3D) and auditory, other health sensing |
| **S7.4** | **Algorithms and architectures for energy efficient motor control / drive** |
| **S7.5** | **Structural health monitoring applications for infrastructure including bridges, buildings, pipes** |
| S7.6 | Algorithms for non-invasive sensing |
| **S8** | **SW / HW Co-development** |
| *SW/HW co-development includes environment and techniques to design, optimize and validate system function, performance and security across the SW/HW boundary. Hardware assistance for runtime and OS functions can improve interrupt response time, provide very-low-jitter compute and I/O, and reduce overhead. Operating system integration with and control of hardware can be used to improve performance (latency, throughput) and power consumption (example: OS could change HW cache management algorithms based on the scheduled workload).* | |
| S8.1 | Optimizing SW-HW interfaces |
| S8.2 | Communications of information between OS/run time/application and the HW platform |
| S8.3 | OS/run time control of HW operation (e.g. choose optimal HW caching algorithm, thread execution priority, etc) |
| **S8.4** | **Hardware schedulers and hardware acceleration for runtime and OS functionality, from real time to high performance, such as for reduced interrupt / event response time, jitter and scheduling and synchronization overhead.** |
| **S9** | **Security and Anti-Counterfeiting** |
| *While Security and Anti-Counterfeiting are important concerns to the SLD thrust, they are central focuses of the Trustworthy and Secure Semiconductors and Systems ("T3S") thrust.* ***Research primarily relevant to S9 must be submitted to the current or next-upcoming T3S solicitation rather than this solicitation.*** *However, papers primarily focused on S1-S8 will be considerably strengthened if they also address topics presented in S9.*<br><br>*Security concerns and counterfeiting are increasingly undermining the hard-won reputations of SoC manufacturers and their customers, and sometimes even compromising bodily safety. Formal analysis is emerging to join simulation-based techniques for detection of attack vectors and estimation of their risk levels. Proposals must involve Security/anti-counterfeiting mechanisms that have hardware-design ramifications. Example application areas may include but are not limited to authentication, smart grid, automotive, medical patient data access, medical implants, point of sales, and separation of operational from public/general-use networks (e.g., preventing infotainment networks from becoming an attack vector into Automotive or CAN based operational networks).* | |
| S9.1 | Formal verification of security |
| S9.2 | Security breach detection, mitigation, and recovery techniques |
| S9.3 | Methods and tools for measuring/checking system security risk |
| S9.4 | Mitigating safety risk due to tampering with safety-critical systems |
| S9.5 | Anti-counterfeiting approaches utilizing/building on PUF and/or non-PUF techniques |