Co-designing from Atoms to Architectures

A. Aiken (Stanford), J. Baniecki (SLAC), J. Bokor (Berkeley), V. Churavy (MIT), R. Coffee (SLAC), A. Dragone (SLAC), A. Edelman (MIT), D. Ernst (Cray), S. Hearne (ORNL), J. Hennessy (Stanford), H. Mabuchi (Stanford), P. McIntyre (Stanford & SLAC), N.A. Melosh (Stanford), Sh. Misra (Sandia), S. Mitra (Stanford), J. Neaton (Berkeley & LBNL), T. Ogitsu (LLNL), O. Ovchinnikova (ORNL), J. Rabaey (Berkeley), A. Salleo (Stanford), E. Schwegler, (LLNL), M. Shulaker (MIT), R. Stevens (U of Chicago & Argonne), T.P. Straatsma, (ORNL), J.B. Thayer (SLAC), N. Wichmann (Cray), H. P. Wong (Stanford), R. N. Zare (Stanford), and S. Shankar[1] (Harvard)

1. Topic: The following document is a Position Paper on Co-design as a methodology for integration of application-specific computing to solve real challenges. This methodology is based on concurrent design and optimization of the computing architecture, software, algorithms, materials, devices, information processing, communication, etc. Previous analysis of specific examples of Co-design (e.g., Anton for protein folding[2]) indicates great promise for acceleration of computing for applications on top of Moore's law. However, these examples are still a selective few with dedicated resources and budgets, and no current framework exists to more broadly integrate the various components of computing for most applications. We believe that a radical rethinking of innovation across components of computing would be useful to address both the multi-disciplinary complexity of constructing new application-specific computing systems and the continuing offshoring of the US innovation ecosystem[3].

2. Co-design Scientific and Technological Challenges: Development of new design abstractions as an integral part of the codesign process remains a fundamental methodological challenge. On a more focused level, low-level aspects of codesign could be greatly accelerated by the development of cyber-physical simulation toolkits and workflows that facilitate simple construction of models in which dynamical systems with real-time control algorithms or other rule-based adaptive supervision. High-level aspects of codesign will likely demand new theoretical tools related to algorithm design for complex heterogeneous hardware systems that may exhibit behavioral uncertainties that cannot effectively be captured by low-dimensional parameterization. We see three types of challenges in a general-purpose computing solution for all applications.

Efficiency: Aspects of efficiency include energy per bit, computational complexity of an application, and manufacturability. Of these, energy or power minimization is a universal macro-constraint for on-chip architectures. The computer industry is actively dealing with trade-offs between performance and energy efficiency. Detection and real-time mitigation of manufacturing abnormalities in the lower levels (materials/devices/circuits) of the system stack is crucial for maintaining system performance, yield, and reducing waste and energy consumption in the manufacturing process. In addition, trade-offs between large volume manufacturing of a few systems versus custom-manufacturing for multiple systems need to be systematically evaluated. Although there are many efforts to analyze energy efficiency and complexity, no dedicated efforts exist that integrate efficiency for researchers and scientists to design and develop tools and methodologies for developing the building blocks for an optimal design.

Prototyping and Manufacturing: There are many challenges including the time lag from design to prototyping, ability for integration of novel materials across multiple technologies, co-optimization between packaging and silicon, ability to ramp up production with well-understood cadences, manufacturing within the US for security and resilience. Current approaches to manufacturing are based on existing designs which are optimized for a given process technology. The process from design to product is still largely sequential with many iterations that are becoming increasingly necessary for efficient optimization across performance, cost, and adaptability. Understanding the opportunities and limitations presented by new ideas in materials, devices and circuits has been limited by the practical limitation of testing those ideas at the system level, at least, as a prototype.

Security and Resiliency: Even if the above building blocks are available for a wider audience, security of the components and the integrated systems are key to resilient computing systems. This aspect is critical for both the components of co-design and the tools used to design, prototype and manufacture. The need for trust in microelectronics design along with implementation of software packages is a critical need for all sectors of computing and system control. The problem is further compounded when sensors designed and manufactured by third party suppliers are integrated into large-scale computational systems. The security and resiliency at the system and component levels are necessary for mission critical applications.

---

[1] *Contact for correspondence*

[2] Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, et al. (2008); "Anton, a special-purpose machine for molecular dynamics simulation". Commun ACM 51:91–97

[3] Co-design Meetings, Harvard (2017); Stanford/SLAC (2019)

3. Guiding Principles for Co-design: A new proposed framework should enable a modular, "Legos-type" approach where researchers could simulate and put together the building blocks to design and prototype the computing advances as they become available. The components themselves should be based on all aspects of efficiency. We propose a few guiding principles below related to global co-design for optimizing efficiency, use of smart hardware-software labs for ability to emulate and prototype, open workflows and tools to design for resilience and testing, and a hub for a multi-disciplinary engagement to engage in cross-cutting research.

Use of Global Co-design from Atoms to Architectures (Bottom-up Design): As proposed, Co-design must reach all the way down to the atomic level to integrate materials engineering and advanced characterization early in the process. The rise of multifunctional atomically engineered materials opens up unprecedented opportunities for bottom-up engineering of building blocks for computing, starting from single atoms or even single electrons. Due to its engineered functionality, the material or the molecule thus becomes the device, which is to be reproduced and integrated in a circuit and, later, a system. This design approach in turn will help integrate information processing between different classes or materials to systems such as analogue, classical, and quantum systems.

Design of Intelligent Hardware-Software Labs: Integration of Artificial Intelligence/Machine Learning (AI/ML) with Hardware/Lab Equipment can enable smart design, development, testing, and prototyping.[4] The current proliferation of opportunistic applications of AI/ML tools in scientific workflows are short-circuiting the detailed physics and chemistry-based analysis of critical complex phenomena. Developing powerful yet human-interpretable AI/ML is still a grand challenge for the AI/ML field in general, but this challenge of overlaying principled statistical theory with practical "big data" can be enabled by Co-design. The concepts and methods from cybernetics and robust control theory and from the emerging field of control-over-channels, could prove useful if properly translated to the codesign context.

Develop Open Workflows and Tools: The exponential increase in sensors both in high-energy physics detectors and in Internet of Things, especially in large-scale scientific facilities and in intelligent consumer systems, pre-supposes a commensurate revolution in data analysis workflow. Realizing this revolution requires multiple stages to transparently handle the acquisition, processing, transfer, analysis, and visualization. Optimization of the workflow to extract information that distills data into actionable information on the timescales required of experiments requires co-design of detectors, edge computing layers, and data analytics which may run on a variety of architectures, from local compute clusters to Department of Energy's leadership computing facilities.

Create Application-Enabled Innovation Hubs: Development of an "innovation hub" will enable the community to use the building blocks for developing physical computing prototypes for visualizing and testing new designs. We propose a two-dimensional approach: 1) Scientific and Engineering Research linking Applications, Architectures, System and Devices, Novel materials along with their synthesis and processing, Information Abstractions; 2) Prototyping for exploring various computing options by developing following components as needed: Design methodologies, Validation strategies, Tool sets for design, Fabrication, Integration, and Packaging. The sensing may be precise in particle accelerators or noisy in wireless networks. Working across the stack in applications, systems software and hardware creates opportunity for innovative Co-design and enabling a disciplined process to bridge/apply shared knowledge across the spectrum.

4. Timeliness of Co-design: With the increasing difficulty of sustaining the cadence of Moore's Law, there is a time critical need to rethink how can we build the next generation of computing while lowering the increasing costs of design and manufacturing. Our vision will enable a new era in personalized and application-centric computing ("Cambrian" era[5]) that bridges information theory, computing and communication abstractions with materials, devices, hardware, systems, architecture, algorithms and software for enabling new applications. As early examples have shown the promise of this approach, our integration would enable a new systematic approach for designing and building efficient computing systems from atoms to materials to devices to systems, which can be rapidly prototyped within an innovation hub.

---

[4] Dally, W. et al. Hardware-enabled artificial intelligence. In Proceedings of the Symposia on VLSI Technology and Circuits (Honolulu, HI, June 18–22). IEEE Press, 2018, 3–6.

[5]Hennessy, J. and Patterson, D. A New Golden Age for Computer Architecture. Communications of the ACM, Vol. 62, No. 2, February 2019.

Supplementary Materials: Co-design from Atoms to Architectures

A. Aiken (Stanford), J. Baniecki (SLAC), J. Bokor (Berkeley), V. Churavy (MIT), R. Coffee (SLAC), A. Dragone (SLAC), A. Edelman (MIT), D. Ernst (Cray), S. Hearne (ORNL), J. Hennessy (Stanford), H. Mabuchi (Stanford), P. McIntyre (Stanford & SLAC), N.A. Melosh (Stanford), Sh. Misra (Sandia), S. Mitra (Stanford), J. Neaton (Berkeley & LBNL), T. Ogitsu (LLNL), O. Ovchinnikova (ORNL), J. Rabaey (Berkeley), A. Salleo (Stanford), E. Schwegler, (LLNL), M. Shulaker (MIT), R. Stevens (U of Chicago & Argonne), T.P. Straatsma, (ORNL), J.B. Thayer (SLAC), N. Wichmann (Cray), H. P. Wong (Stanford), R. N. Zare (Stanford), and S. Shankar (Harvard)

APPENDIX 1: *Comments on Areas of Emphasis*

APPENDIX 2: *Answers to a few Notational Questions*

APPENDIX 1: *Comments on Areas of Emphasis*

1. **Key aspects of codesign across the entire hardware/software stack to include applications, algorithms, system software, and system architecture;**
    - Novel approach to co-design is needed, that bridges from applications all the way to materials. Traditional co-design has focused on tradeoffs between algorithms, (software) and hardware implementations of functions. Permitting trades in power, performance, flexibility, cost, etc. In the future we need to consider deep trades due to additional alternatives such as surrogate models of applications, broader options in architecture and potentially novel materials that permit improvements in power.

2. **Insights into codesign for workflows arising from scientific experiments, on supercomputers, or to support large-scale scientific instrument;**
    - Pushing past traditional single applications to entire end to end workflows implies that you might have different systems components carrying out different stages of computation and I/O. This opens up the idea of hierarchical or multi-component optimization as part of the co-design with different trade offs being exploited for different elements of the workflows.

3. **Methods and tools for quantitative codesign, including both those that inform high-level decision-making and those impacting low-level aspects of the codesign process;**
    - Design synthesis (either traditional or via new AI driven methods) will be an important future driver of co-design. The idea of extending design synthesis to the full stack should be considered as it may result in deeper optimizations. Iterative design is another approach that revisits design choices multiple times as different trajectories are explored.
    - Principled approaches to the development of new design abstractions as an integral part of the codesign process remains a fundamental methodological challenge. On a more focused level, low-level aspects of codesign could be greatly accelerated by the development of cyber-physical simulation toolkits that facilitate simple construction of models in which dynamical systems (specified in terms of ordinary or partial differential equations) interact in real time with control algorithms or other rule-based adaptive supervision. High-level aspects of codesign will likely demand new theoretical tools related to algorithm design for complex heterogeneous hardware systems that may exhibit behavioral uncertainties that cannot effectively be captured by low-dimensional parameterization. It seems likely that concepts and methods from robust control theory (and from the emerging field of control-over-channels, e.g., Bode-Shannon theory) could prove useful in this context if properly translated to the codesign context.

4. **New codesign challenges anticipated over the next decade**
    - The dream of single computational systems where analog and digital computation is mixed in to a hybrid computation system has been simmering in the background of computing for decades [A. Hausner, Analog and analog/hybrid computer programming, Prentice-Hall, 1971]. The recent advances in quantum computing and integrated non-linear photonic systems [N Singh, et. al. Photonics (2020) https://doi.org/10.1364/PRJ.400057] have greatly expanded the potential capabilities and applications of these systems. However, the complexity of building, optimizing and implementing such computational systems has stymied their applications to anything more than a few niche applications. This is an opportunity for co-design of materials, devices, and system architecture provide a unique opportunity to greatly expand the impact of hybrid computing. Challenges include understanding dissimilar interfaces, and joining of analog and digital components together, as well electrically / optically coupling the devices.
    - From an applications perspective, the next decade will see dramatic new challenges in public health that can only be met by revolutionary codesign methodologies. Tracking the emergence and mutation

of novel viruses will require widespread environmental sampling and genomic/proteomic characterization on unprecedented scale; sheer numbers will require innovative codesign of hardware systems and search algorithms in order to enable surveillance at scale within achievable supply rates of biochemical reagents, dedicated laboratory time, and compute cycles. The next decade will likewise see a dramatic increase in the number of autonomous/auto-piloted vehicles on roadways and in commercial airspace; robust, secure, verifiable and adaptable/updatable collision avoidance systems will likely require codesign of sensors, transponders and algorithms/protocols at their core. From a methodological perspective, the generalization of nascent codesign principles beyond the relatively familiar substrate of semiconductor electronics to biological and quantum systems will demand accelerated development of core theory regarding codesign as a new science in its own right.

- System on a Chip and flexible hardware design to accommodate variety of the needs in industry and scientific community. It is apparent that there are only a handful of companies who are capable of mass producing SOC with sufficient size of on chip high-bandwidth memory in economical way. Most likely, HPC project will also rely on their capability. Designing capability of a prototype is important but supplying sufficient number of HPC chips is critical.

APPENDIX 2: *Answers to a few Notational Questions*

1. **How to balance breadth of applications versus customization benefit? Does this vary by system type?**
    - This is a hard question to answer in general for the breadth of applications. For it to be viable, the breadth of applications includes enough users to put the project above the threshold to recover the fixed costs of designing and building hardware. And to the extent that we can make codesign more efficient we help reduce those fixed costs and the risk associated with any new codesign project. It most certainly will vary by system type and use case, but by focusing on those use cases with higher potential gains (that could be cumulative) will keep things moving. With some scenarios that are exploiting surrogatges we can see factors of thousands in improvement coming from alternative algorithms with changes to hardware needed to support those new methods.
    - System on a Chip and flexible hardware design to accommodate variety of the needs in industry and scientific community. Designing capability of a prototype is important and needs to be connected to scaled-up manufacturing.
    - The resent revolution in additive manufacturing has been fired by the ability to use a library of materials to produce nearly an infinite number of new structural that solve specific problem. Co-design enables this same revolution to occur in microelectronics manufacturing.

2. **What are the tools and techniques that enable successful codesign interactions and where are the gaps?**
    - We need next generation of vertically integrated design and simulation tools and high speed emulators to run them on.

3. **With artificial intelligence and machine learning becoming more widely used in scientific workflows and applications, what new challenges and opportunities does this present?**
    - The current proliferation of opportunistic applications of "black box" artificial intelligence and machine learning (AI/ML) tools in scientific workflows has the unfortunate consequence of short-circuiting the traditional emphasis on painstaking careful development of new concepts and reduced models for critical complex phenomena. As a result, hand-waving interpretations of otherwise inscrutable designs discovered by application of AI/ML to limited datasets launch dubious ideas into complex systems engineering, which may in fact reflect only improper generalization of specific results to broader context, or even unwarranted rationalizations of overfitting by AI/ML routines. Of course, developing powerful yet human-interpretable AI/ML is a grand challenge for the AI/ML field in general, but this challenge of overlaying principled statistical theory with practical "big data" seems to echo the broader paradigm of codesign. There could be opportunities to help guide the development of AI/ML practice within the mainstream scientific community by establishing "governed" cloud computing services that incorporate not only data storage and computation but rigorous meta-analysis of results and curated guidance regarding best practices for interpretation of specific types of results.
    - The primary one is that the low level architectural features needed to support AI centric workloads are different than traditional architectures and in many ways are simpler and need to be part of the design space considerations.

4. **New accelerator technologies and chiplets are increasing the possible design space. What is the potential of these new technologies and how can codesign be used to take maximum advantage of them?**
    - Co-design is to provide accurate information regarding application side requirement as well as constraint coming from options in fabrication capabilities thereby most effective platform for developing future HPC chipset that can be the most (cost) effective for both scientific and industry uses. If one admits SOC is the way to go, understanding possible design space will

likely to have constraints coming from the fabrication capability is must. Understanding could be achieved if good communication is established.

- Essentially these open the space of solutions a bit by making it easier to integrate technologies from different nodes etc. A major breakthrough would be possible if we could use these new packaging technologies to integrate components fabricated with different materials. Also one could consider that we now have the ability to map IP to chiplets and do fine grain integration.

5. **How can we further the state of the art in efficient and flexible open-source hardware, modeling, and simulation tools that can underpin hardware codesign activities?**
    - Co-design of materials, devices, and architecture dramatically increases the parameter space for engineering the next generation of computational systems. It will only be possible to take full advantage of these advances if there are models that can rapidly explore the parameter space in real time during fabrication.
    - Encourage open source design stack with well defined APIs that encourage groups to develop components. Require that for government procurements that in addition to the hardware that the vendors ship the full simulation stack so it can be used as baselines for further co-design work.

6. **Is there a performance benefit to codesigning scientific applications and the computer systems (hardware and software) they run on, or will the additional time and cost outweigh the benefits observed relative to more-or-less portable applications running on stock supercomputers?**
    - A significant benefit to the additional complexity of co-designed hybrid systems is the ability to improve the digital / materials fingerprint for increased security. The need for trust in microelectronics design along with implementation of software packages is a critical need for all sectors of computing and system control. Co-design greatly increases the ways in which unique digital fingerprints can be related to enable trusted systems.
    - With the characteristics of next generation scientific facilities in National Laboratories and elsewhere, experiments requiring massive-scale data analytics are on the horizon. Co-design of smart sensors, customized computing systems and real-time algorithms, not only will increase performance but would constitute an enabling factor. Co-design will enable a new level of customization and optimization focused on information extraction and data reduction throughout the entire data acquisition chains.

7. **How do scientific applications and supercomputer codesign differ fundamentally from the codesign employed regularly for embedded systems (such as in automobiles and home appliances)? Can either area learn from the other?**

- Fundamentally the ability to simulate end-to-end is much greater in embedded systems and this enables co-design. Addressing the simulation stack is one way to improve this.