SUBVERSIVE ARCHITECTURES

Roadrunner and the Path to Exascale Supercomputing

Author: Nicholas Lewis, University of Minnesota. Mentors: Gary Grider, Carolyn Connor, Jeffrey Johnson.

Los Alamos NATIONAL LABORATORY EST. 1943 — **HPC History of LANL Supercomputing Project**

LA-UR-17-26949

The Road to Roadrunner

Roadrunner in the SCC, 2010. *LANL.*

A collaboration between LANL and IBM, the Roadrunner cluster computer operated between 2008 and 2013, and was the first supercomputer to break the petaflop/s barrier, performing over one-quadrillion sustained floating-point operations per second. The first large-scale "hybrid" cluster, Roadrunner was both unique and a harbinger of what was to come. It was unique in that, even as Roadrunner proved the viability of heterogeneous computing, Roadrunner's particular architectural approach, combining conventional CPUs with Cell accelerator chips, ended with IBM's cancellation of the Cell chip line. Yet, as a harbinger, Roadrunner's development contended directly with some of the most significant issues facing the High-Performance Computing field, such as performance per watt and HPC's declining share of the overall computing market, an issue impacting the technical and supply options underlying new systems. HPC centers, even more so than in the decades before, had to become creative in their selection and application of the products that a limited range of vendors were willing to supply, amid HPC's search for viable paths around its encroaching technical hurdles. Roadrunner's place amid the complex supercomputing environment and its challenges in the early twenty-first century is best understood through the context of other pathfinder machines, like the Cray-1 and the Connection Machines, which introduced new architectures, and brought both obvious and subtle disruptions, to the HPC field. This reveals that Roadrunner's was not the first contested architecture in the HPC community. However, many of its struggles were characteristic of the HPC field following the supercomputer-market crash of the 1990s, the shrinking place of HPC in computer market share in the early 2000s, and the increasing reliance upon simulation for the DOE's Stockpile Stewardship Program.

Subversive Architectures

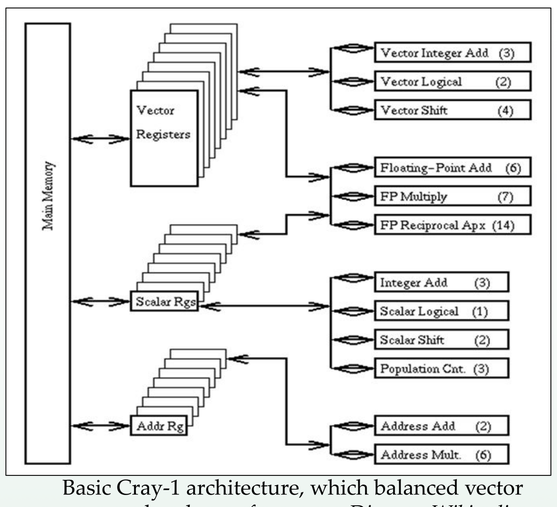
Increases in computing hardware performance have typically been achieved in two ways:

1) Improving component technology (e.g. lower gate delays, faster memories).

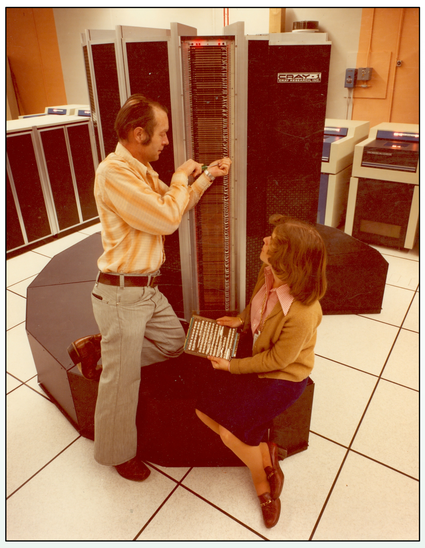
2) Changing computer architectures (e.g. the transition from scalar-only to vector architectures).

Technical roadblocks in the former have typically spurred the latter, with most architectural changes since the 1960s involving the addition of various forms of parallelism or concurrency. New architectural approaches have encountered differing levels of resistance and acceptance in the HPC community, depending on a variety of social, economic, and technical factors.

Examples of these architectural shifts include:



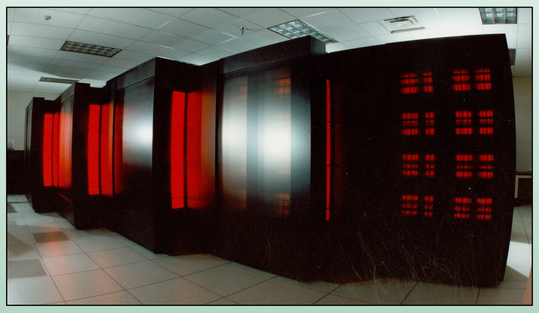
Basic Cray-1 architecture, which balanced vector and scalar performance. *Diagram Wikipedia.*



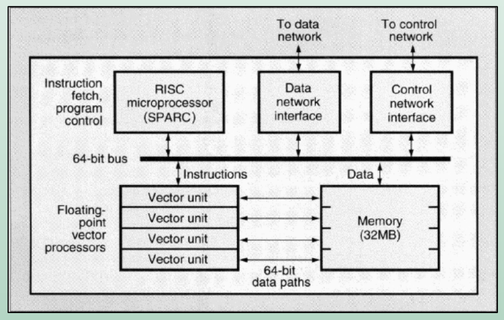
Cray-1 in the CCF, 1978. *LANL.*

Cray-1 (1976)

- The first commercially successful vectorized supercomputer.
- The Cray-1 balanced vector and scalar performance, making the Cray-1 less of a risk of disruption for existing supercomputer users than its vector-oriented competitors.
- Market-share among new supercomputer customers and a growing range of software brought the Cray-style architecture to dominance in the HPC community from the late 70s through the 1990s.



Thinking Machines CM-5, 1994. *LANL.*



Basic CM-5 architecture. *Diagram Wikipedia.*

Thinking Machines Corporation Connection Machines (CM-1, 2, 200, 5)

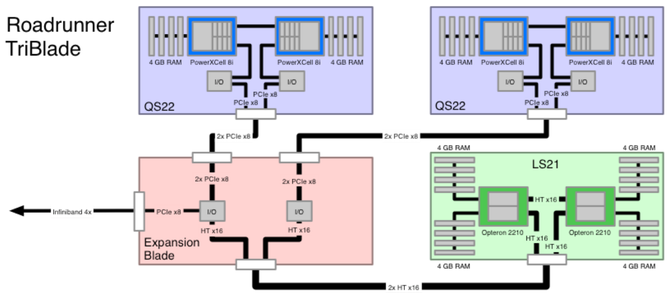
- Among the first commercially viable massively parallel architectures (SIMD; MIMD for CM-5).
- Most TMC (founded in 1983) computers utilized thousands of one-bit microprocessors, which reduced the wastefulness of large word sizes, and permitted more effective parallelization.
- CM range found use in certain scientific and cryptologic applications, but most supercomputer centers had little incentive to transition from vector architectures. Most centers were not encountering the technical limits of vector computing until the later 1990s.

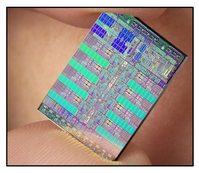
Whence Roadrunner

Following the supercomputer-market crash of the early 1990s, and the disappearance of many traditional supercomputer vendors, cluster computing developed quickly as a means of leveraging inexpensive commodity hardware that had grown out of personal computer development. The Accelerated Strategic Computing Initiative (ASCI, now ASC) was a DOE effort to stimulate cluster advancement in the face of Stockpile Stewardship's demands on supercomputer simulation. A large part of the economic reality behind cluster computing was the diminishing size of HPC in the overall computing market, reducing the market presence and influence of HPC customers compared with previous decades. By the early 2000s, several large ASCI systems, like Blue Mountain, had been delivered to the DOE labs. While cluster performance was increasing rapidly, major hurdles were on the horizon for ASCI's (and the HPC community's) long-term performance goals of pushing beyond the petascale to the exscale over the following two decades. Cluster computing had become the dominant, accepted path to building large supercomputers, but clusters intensified many long-standing problems for the HPC field, while introducing problems of their own.

LANL's ASCI Blue Mountain [left] was the realization of the larger HPC effort to leverage technologies developed for consumer products. Roadrunner's use of a modified version of the PS3's Cell chip extended that effort into the new heterogeneous-cluster paradigm. *PS3 image Wikipedia*

The Harbinger



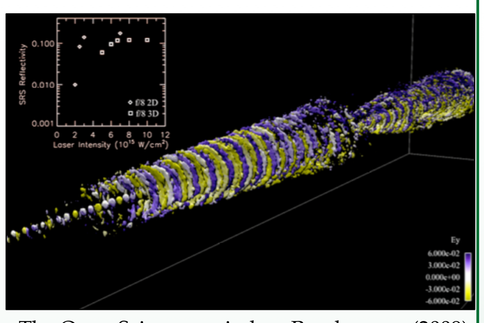


The Roadrunner "hybrid" architecture [above left] was based on an IBM "TriBlade" [above right] (3,240 total) consisting of two, dual-core AMD Xeon CPUs (6,480 total) connected via an "expansion blade" to two additional blades, each equipped with two PowerXCell 8i Accelerator chips [right]. 12,960 total Cell chips were responsible for about 97% of Roadrunner's performance, making Roadrunner an "unbalanced" machine. *Diagram Wikipedia, TriBlade image LANL, Cell image IBM.*

The Roadrunner Project, underway by 2002, started earlier than anticipated, due to funding arriving about two years early, resulting in fewer features being added, and the technologies available being less mature than expected.

The New Architecture

- Major purpose of Roadrunner's new architecture was to mitigate inherent technical limitations of traditional architectures.
- Performance per watt would not scale indefinitely with CPU-only cluster architectures.
- Roadrunner coupled an AMD Xeon general-purpose CPU with special-purpose Cell accelerators.
- Cell chips were fast at short vectors, while being energy efficient, but were otherwise limited.
- Beyond hardware, the new architecture was also intended to disrupt existing programming models.



The Open Science period on Roadrunner (2009) demonstrated that the initial, difficult reprogramming effort had rewards for scientists, like with this record-breaking 400-billion particle simulation of laser-light scattering. *LANL*

The Consequences

- Roadrunner demonstrated that large-scale hybrid clusters could be designed within the constraints of what vendors were able and willing to offer. Vendors like NVIDIA began offering other solutions for hybrid systems.
- Demonstrated that programming hybrids, though difficult, typically paid off for users and application developers. The point was to push programmers to make a difficult change.
- Roadrunner was fast and energy efficient for its time (2.2MW per petaflop vs. 6.6MW per petaflop for Oak Ridge's non-hybrid Jaguar cluster).
- Some lessons learned with Roadrunner, such as interdisciplinarity within programming groups, were applied inconsistently. Other lessons are now being rediscovered independently elsewhere.
- Hybrid clusters now constitute a disproportionate number of fastest machines, and a variety of hybrid-cluster hardware and software solutions are available from vendors on all performance levels. Hybrids are now accepted as an important path to exascale.

Rack-Mounted Roadrunner TriBlades *LANL.*

Conclusion: While computers like the Cray-1 and the Connection Machines appear to have stirred more clear-cut changes to HPC, they arose at a time when a more engaged computer vendor community sought solutions to issues in supercomputing. Roadrunner was the product of a radically different environment, where HPC users and centers assumed much more of the burden for recognizing and validating pathways to the future of HPC. Roadrunner proved the viability and potential of hybrid computing, and was a demonstration of how the HPC community could continue to innovate amid early twenty-first century constraints, without shying away from the difficult choices and work that ultimately laid the foundation for an emerging exascale era.