FAWN

A Fast Array of Wimpy Nodes

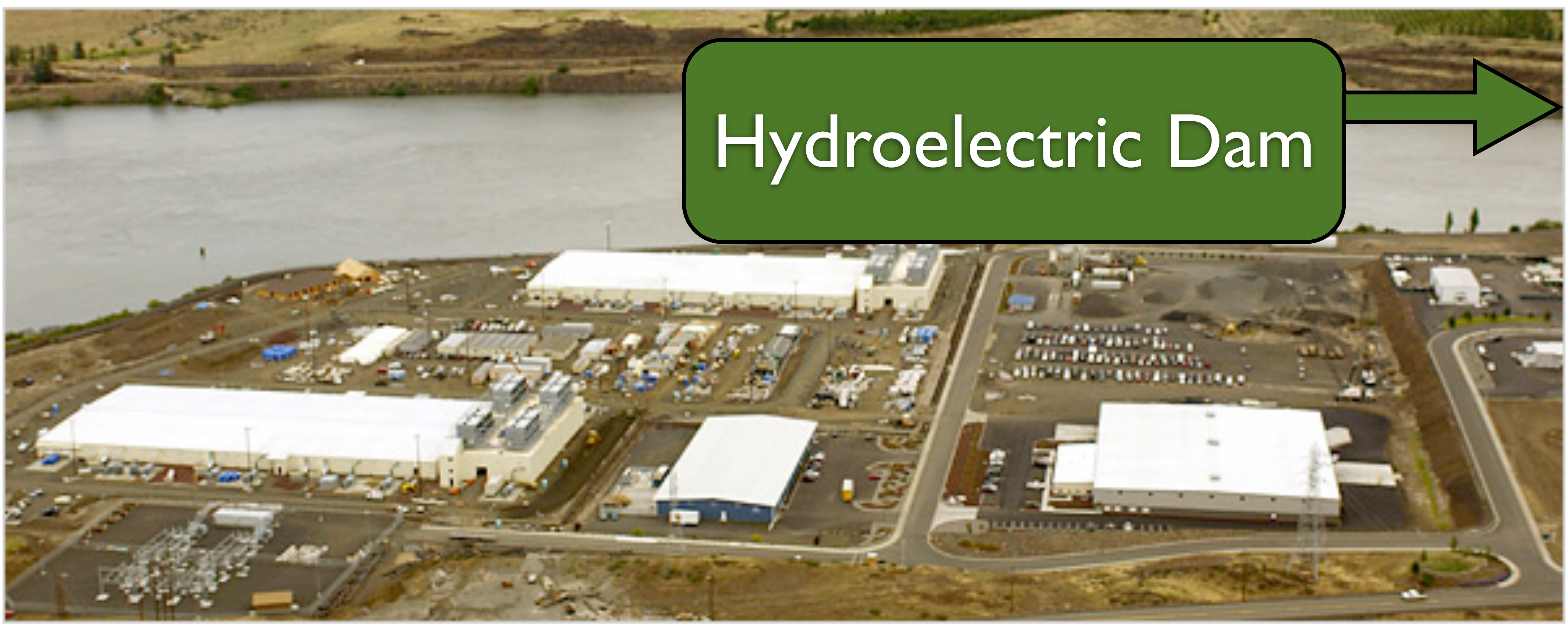
David Andersen, Jason Franklin, Michael Kaminsky*,
Amar Phanishayee, Lawrence Tan, **Vijay Vasudevan**

Carnegie Mellon University *Intel Labs Pittsburgh

October 2009

Energy in computing

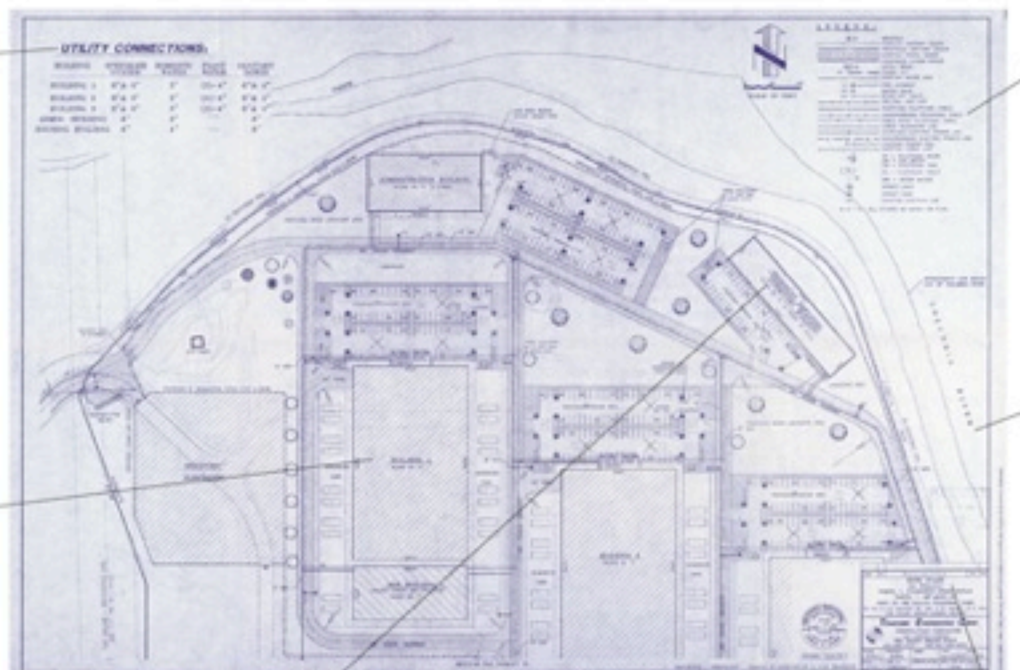
- Power is a significant burden on computing
  - 3-year TCO soon to be dominated by power



KEYWORD: EVIL

Google's addiction to cheap electricity, *by Ginger Strand*

"Don't be evil," the motto of Google, is tailored to the popular image of the company—and the information economy itself—as a clean, green twenty-first-century antidote to the toxic excesses of the past century's industries. The firm's plan to develop a gigawatt of new renewable energy recently caused a blip in its stock price and was greeted by the press as a curious act of benevolence. But the move is part of a campaign to compensate for the company's own excesses, which can be observed on the banks of the Columbia River, where Google and its rivals are raising server farms to tap into some of the cheapest electricity in North America. The blueprints depicting Google's data center at The Dalles, Oregon, are proof that the Web is no ethereal store of ideas, shimmering over our heads like the aurora borealis. It is a new heavy industry, an energy glutton that is only growing hungrier.



Every time someone clicks the "Google Search" button, thousands of servers, like those Google will amass inside these three projected 68,680-square-foot storage buildings, reel into action. (Only two of the buildings have been constructed so far; the company is tight-lipped about how many servers it owns, but current estimates run as high as a million.) A query for *American Idol*, the top search on Google News in 2007, trolls through petabytes of data, using tens of billions of CPU cycles. Velcroed together, stacked in racks, and lined up in back-to-back rows, the servers require a half-watt in cooling for every watt they use in processing, and Google leads the field in squeezing more servers into less space. Based on a projected industry standard of 500 watts per square foot in 2011, the Dalles plant can be expected to demand about 103 megawatts of electricity—enough to power 82,000 homes, or a city the size of Tacoma, Washington.

Google's server farm represents a new phase in the transformation of the Columbia River over the past half-century. Completed in 1957, The Dalles Dam obliterated the area's famous salmon runs by drowning nearby Celilo Falls, a Native American trading site with a peak water volume ten times that of Niagara Falls. The Bonneville Power Administration (BPA), a federal agency that sells electricity from thirty-one dams and one nuclear power plant, then lured aluminum smelters to the region. Across the street from the Google data center is an idle Northwest Aluminum smelter that once used 85 megawatts. In 2000, when energy prices soared, it was decommissioned, and it now is being dismantled for scrap. As the products on which the river's economy depends—fish, metal, bytes—have dematerialized, so has the demand for labor. Northwest and its sibling smelter, Goldendale, employed 1,100 people; Google says it will bring 100 to 200 jobs to the region. And like the vanished salmon, the workers who live in this twenty-unit, faux-rustic transient-employee dorm will merely be passing through.

If any acts of charity figured in Google's arrival at The Dalles, they were the handouts extended to the company by local officials. The real estate deal, announced in February 2005, was delayed six months by Google's conditions—a tax exemption, assurance of cheap energy from the BPA, and the city-built fiber-optic ring indicated here. The state tax breaks and the fiber-optic ring were in place by April, but bargain power could not be guaranteed. With energy prices soaring, the Bush Administration had floated the idea of privatizing the BPA, which would raise the cost of its electricity to market rates. After a conference call between Google, the BPA, and Representative Greg Walden (R., Ore.), the congressman pledged to the press that privatization would be blocked. That August, President Bush signed the Energy Policy Act of 2005, which included an estimated $85 billion in subsidies and tax breaks for the energy business and left the BPA alone. Four days later, Google closed on the land. Thus, through city infrastructure, state givebacks, and federally subsidized power, YouTube is bankrolled by us.

Google's infrastructure buildup has triggered an arms race. Microsoft, Yahoo, and Ask.com are also building data centers on the Columbia River. As they compete to offer software, music, and movies over the Web in the coming era of "cloud computing," they will need more servers running faster and hotter. This way upstream, in Quincy, Washington, Microsoft and Yahoo have contracted for a combined 90 megawatts of electricity—more than the World Trade Center humming at peak power on a hot summer day. The EPA estimates that by 2011, U.S. data-center power use will double, but a quirk in its accounting excluded Google from the study. Even if Google offsets its own energy use with green power or carbon credits, it cannot guarantee that its competitors will do the same. The company's motto is perhaps due for an addendum: "Lead others not into temptation."

In 2006 American data centers consumed more power than American televisions. Google—whose zeal for secrecy is evident here in the data center's code name, O2 PROJECT—and its rivals now head abroad for cheaper, often dirtier power. Microsoft has announced plans for a data center in Siberia, AT&T has built two in Shanghai, and Dublin has attracted Google and Microsoft. In all three locations, as in the United States, the burning of fossil fuels accounts for a majority of the electricity. Google is negotiating for a new site in Lithuania, disingenuously described as being near a hydroelectric dam. But no matter where the data center is located, Google will be tapping into Lithuania's power grid, which is 0.5 percent hydroelectric and 78 percent nuclear. As the functions long performed by personal computers come to be executed at these far-flung data centers, the technology industry has rapturously rebranded the Internet as "the cloud." The metaphor is apt, both for our foggy notions of a green Web and for the storm that awaits a culture that squanders its resources. ■

Ginger Strand is the author of Inventing Niagara: Beauty, Power, and Lies, *to be published this spring by Simon & Schuster.*

ANNOTATION

KEYWORD: EVIL

"Google's power consumption ... would incur an annual electricity bill of nearly $38 million"

[Qureshi:sigcomm09]

"Energy consumption by … data centers could nearly double ... (by 2011) to more than 100 billion kWh, representing a $7.4 billion annual electricity cost"

[EPA Report 2007]

Annual cost of energy for Google, Amazon, Microsoft

=

Annual cost of all first-year CS PhD Students

Can we reduce energy use by a factor of ten?

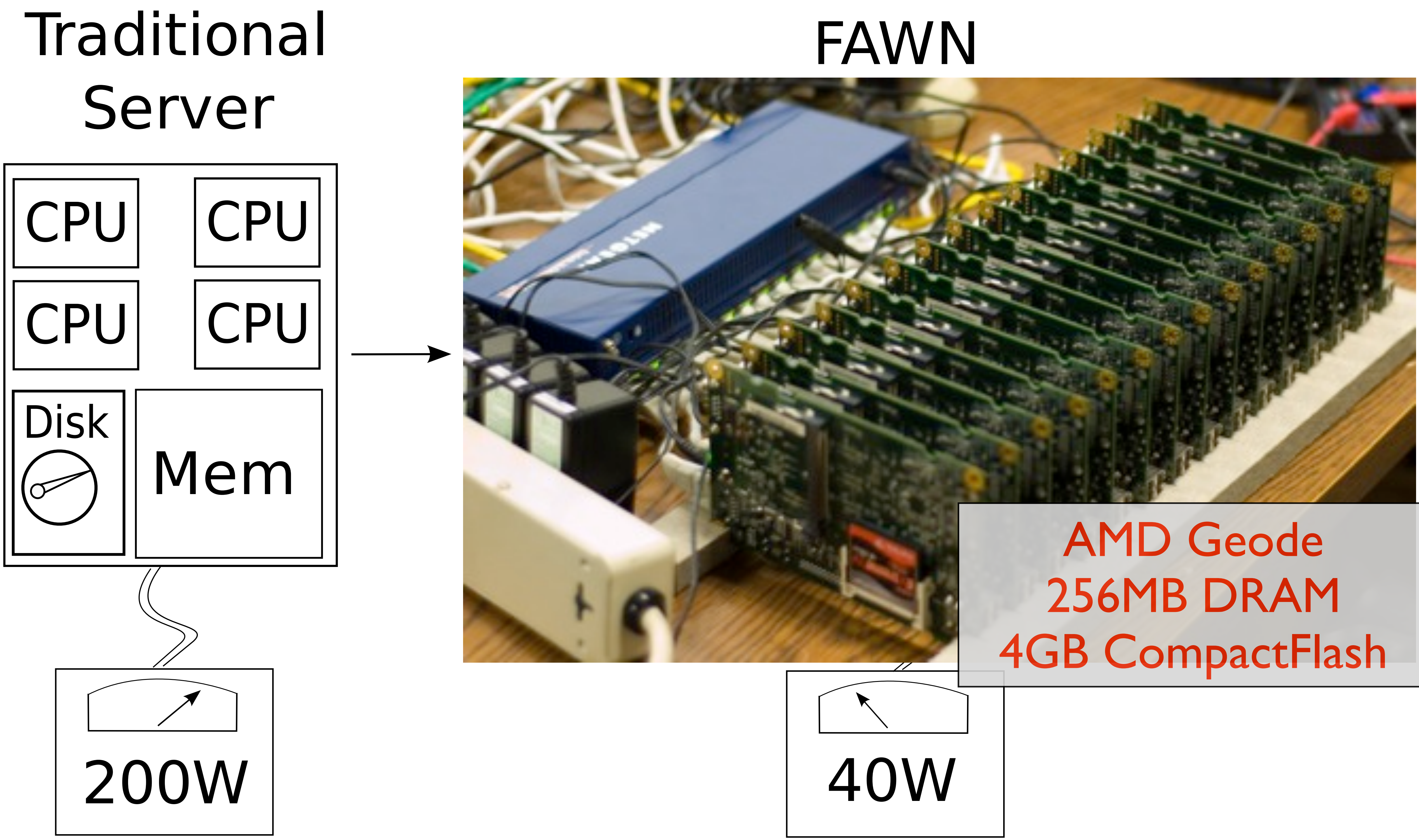
Still serve the same workloads

Avoid increasing capital cost

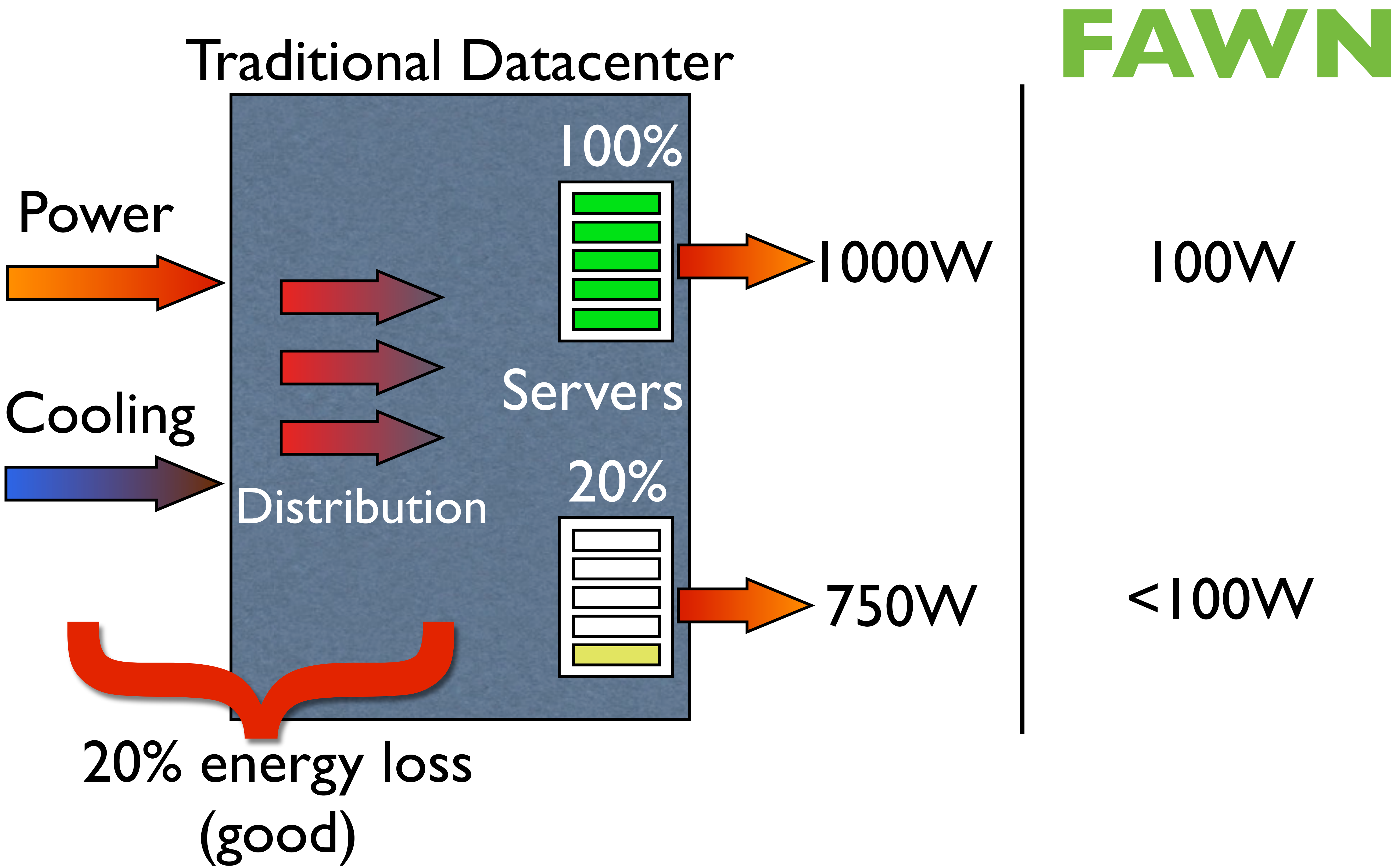
FAWN

Fast Array of Wimpy Nodes

Improve computational efficiency of data-intensive computing using an array of well-balanced low-power systems.



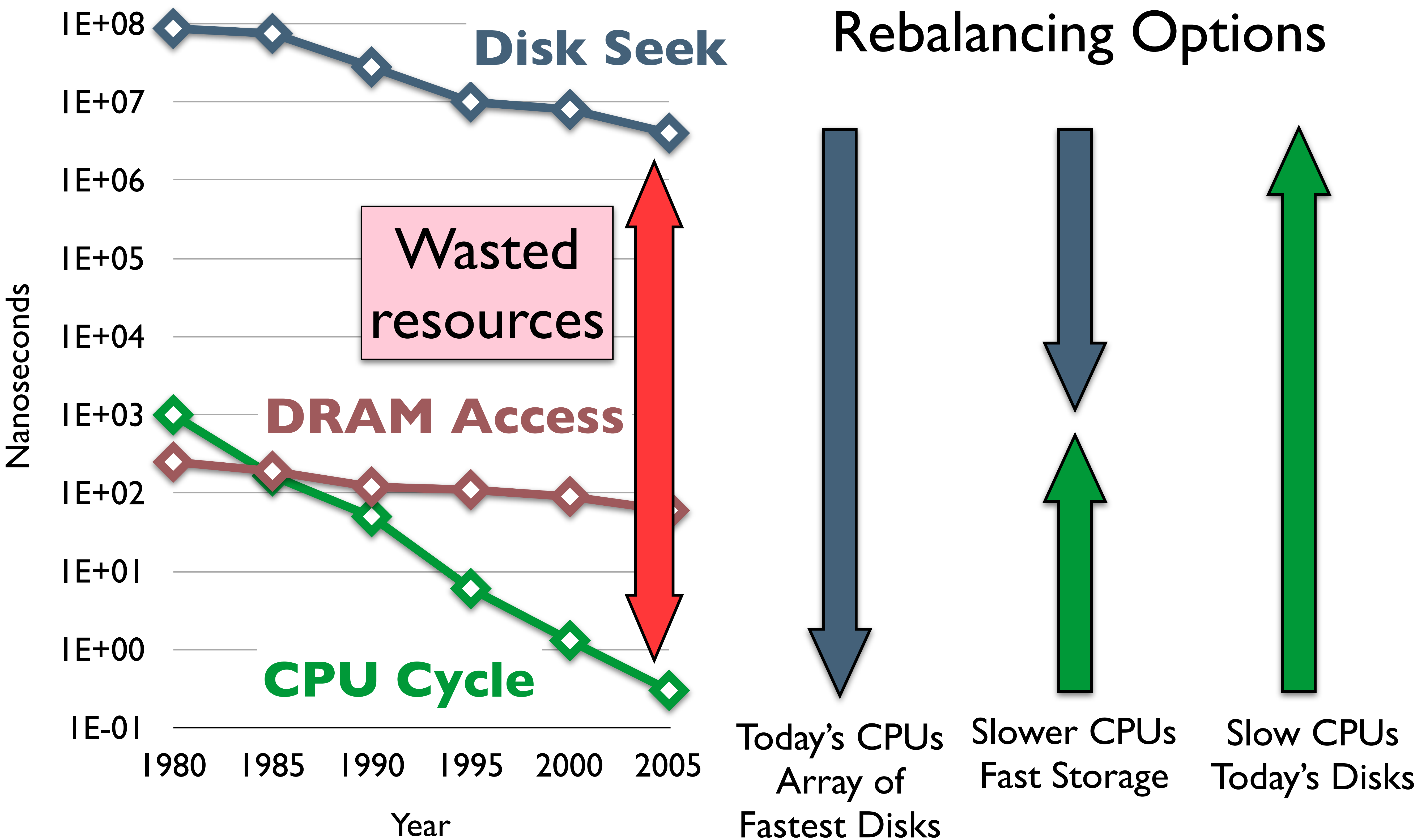
Goal: reduce peak power



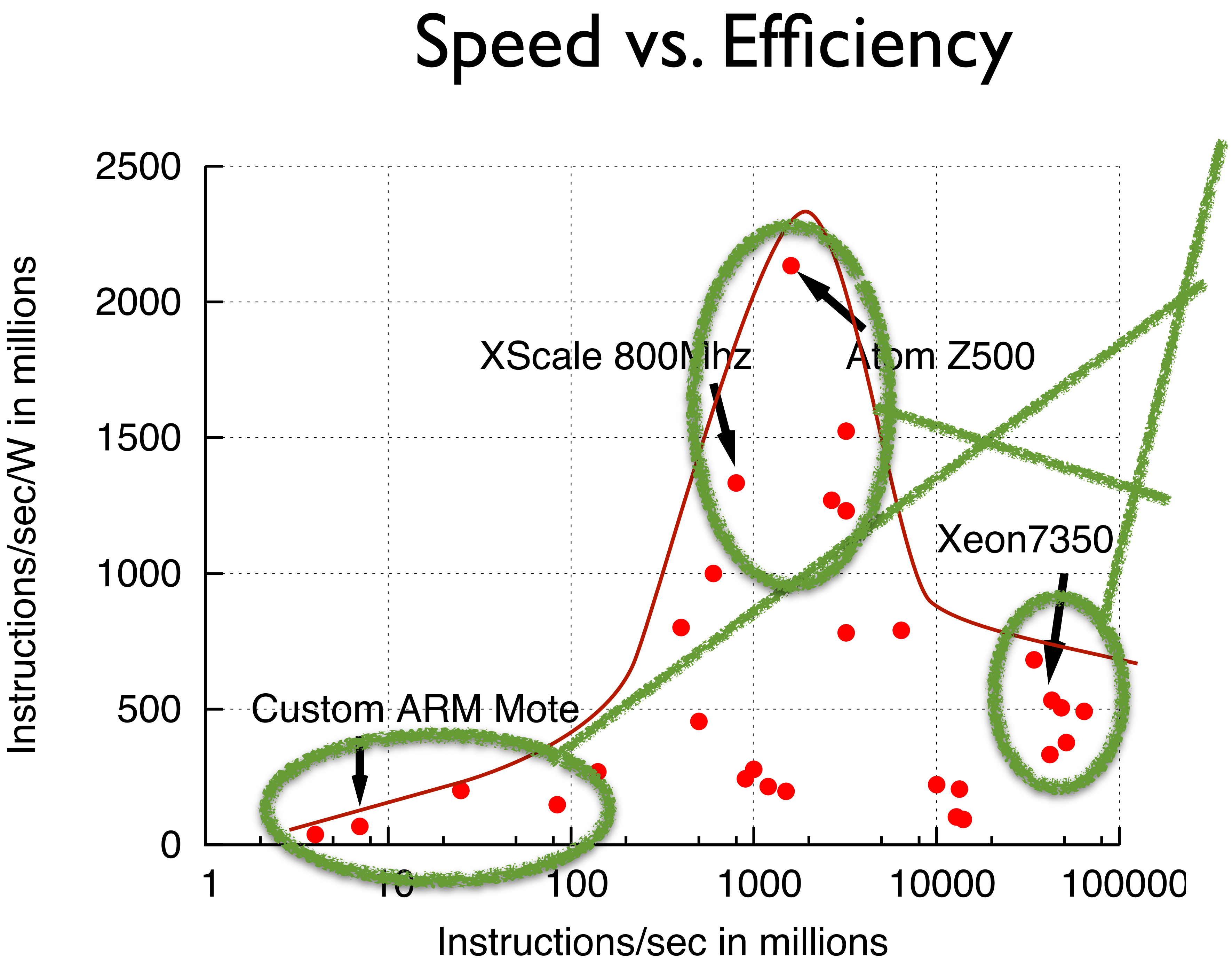
Overview

- Background
- **FAWN Principles**
- FAWN-KV Design
- Evaluation
- Conclusion

Towards balanced systems



Targeting the sweet-spot in efficiency



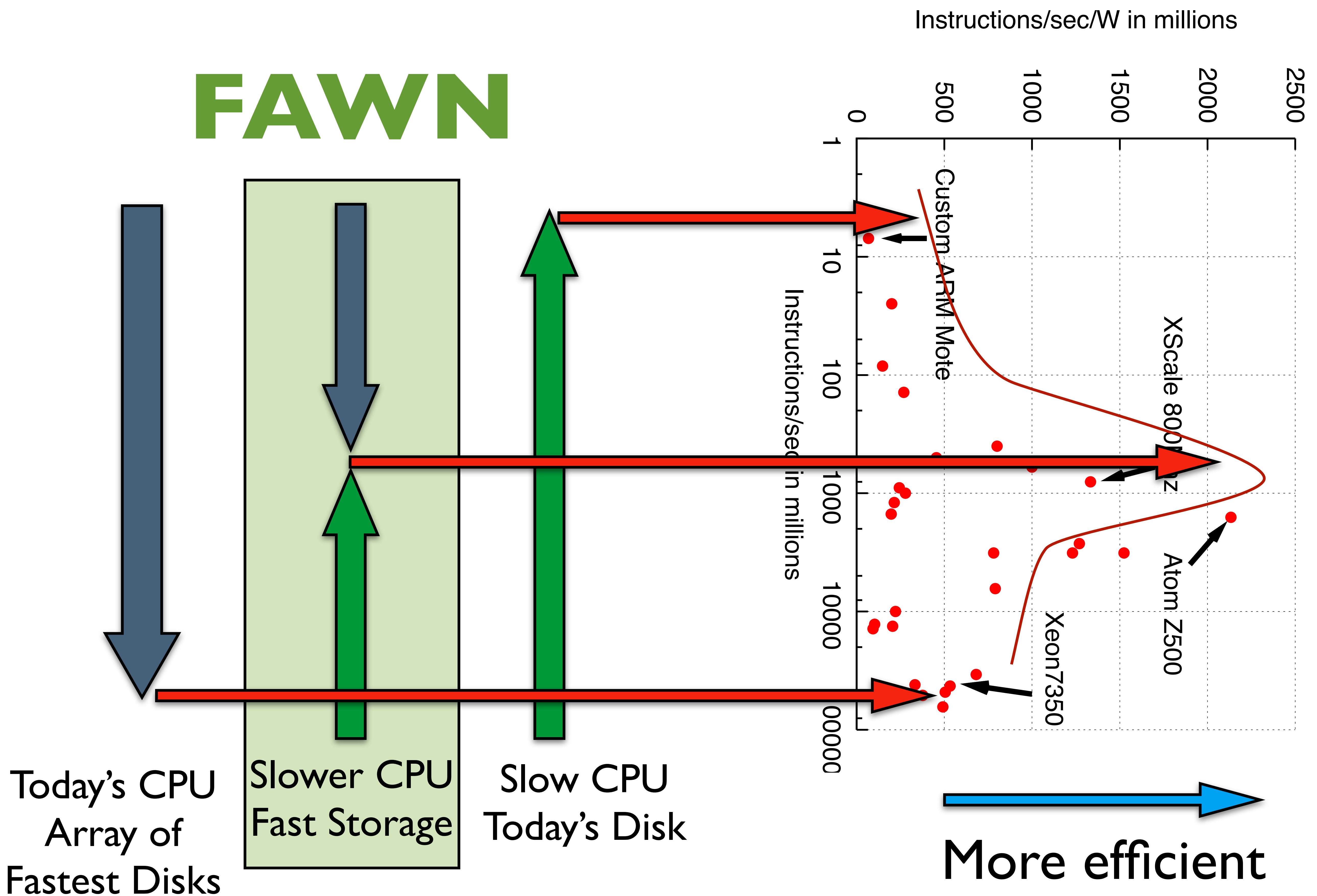
Fastest processors exhibit superlinear power usage

Fixed power costs can dominate efficiency for slow processors

FAWN targets sweet spot in system efficiency when including fixed costs

(Includes 0.1W power overhead)

Targeting the sweet-spot in efficiency

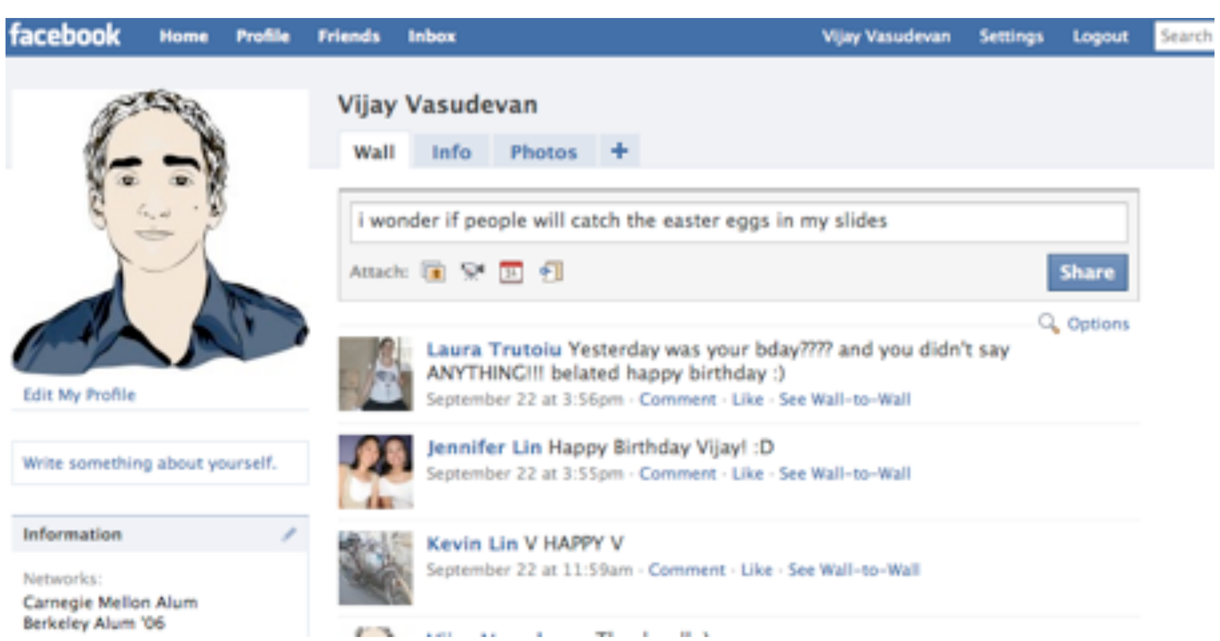


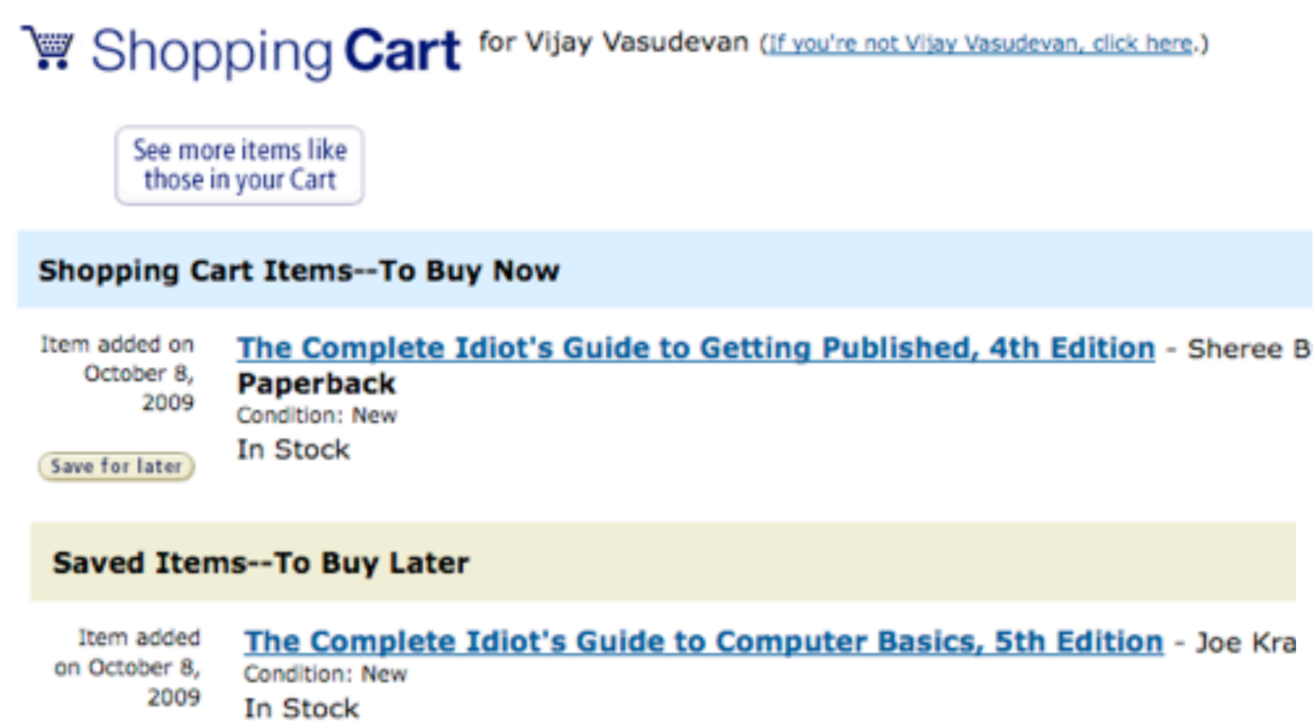
Overview

- Background
- FAWN Principles
- **FAWN-KV Design**
  - Architecture
  - Constraints
- Evaluation
- Conclusion

Data-intensive Key Value

- Critical infrastructure service
- Service level agreements for performance/latency
- Random-access, read-mostly, hard to cache





FAWN-KV:
Our Key Value Proposition

- Energy-efficient cluster key-value store
  - Goal: improve **Queries/Joule**

- Prototype: Alix3c2 nodes with flash storage
  - 500MHz CPU, 256MB DRAM, 4GB CompactFlash

FAWN-KV:
Our Key Value Proposition

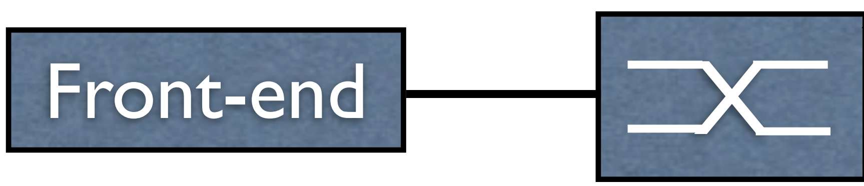
Unique Challenges:

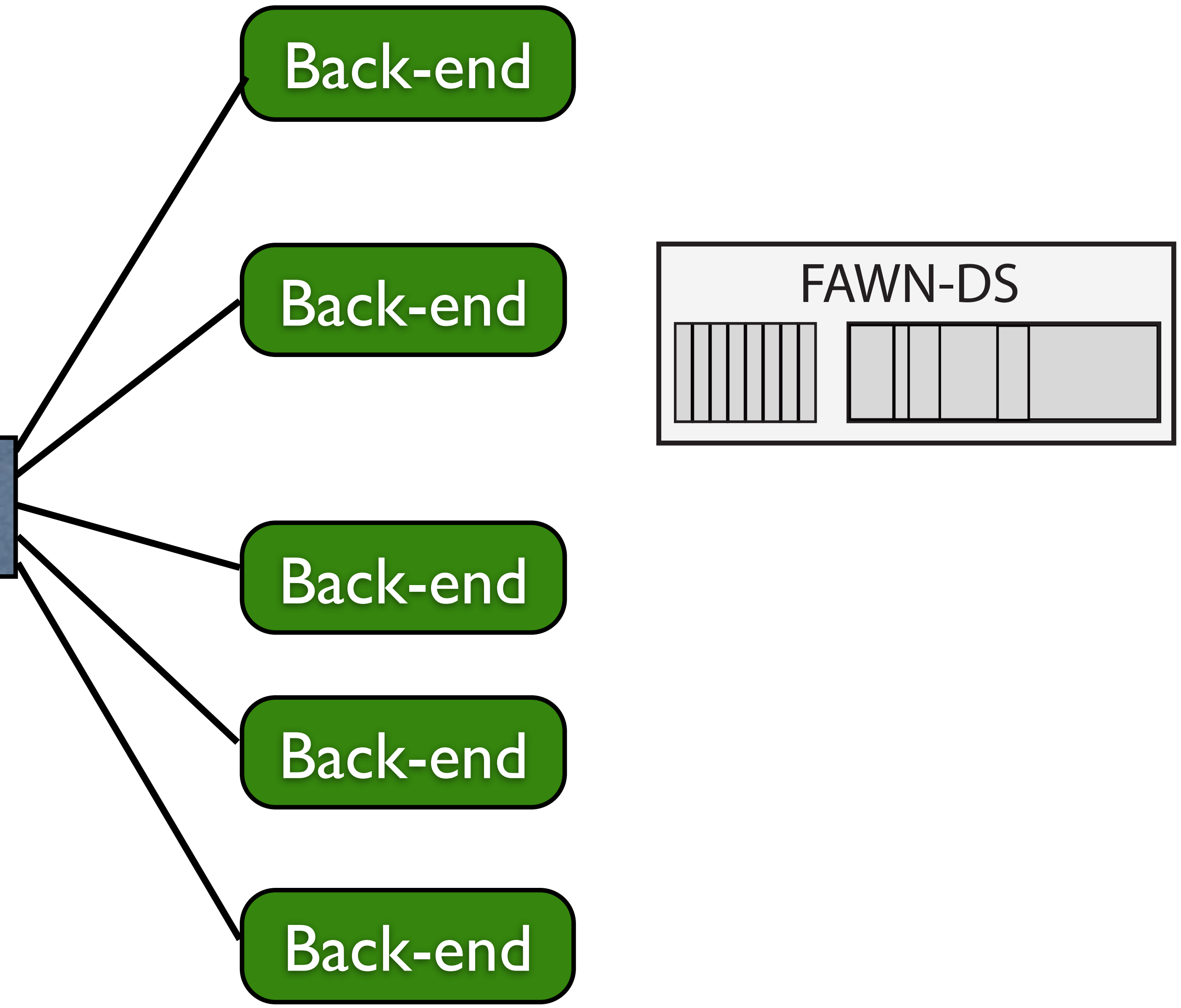
- **Efficient and fast failover**
- **Wimpy CPUs, limited DRAM**
- **Flash poor at small random writes**

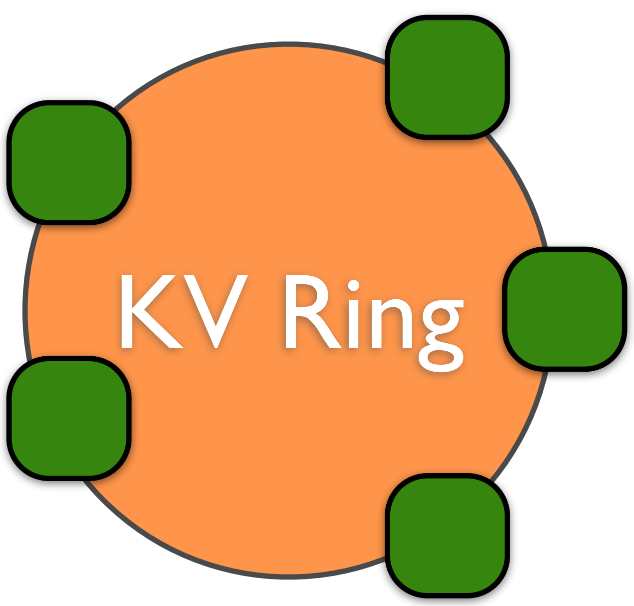
- Prototype: Alix3c2 nodes with flash storage
  - 500MHz CPU, 256MB DRAM, 4GB CompactFlash

FAWN-KV Architecture

Manages Backends
Acts as Gateway
Routes Requests

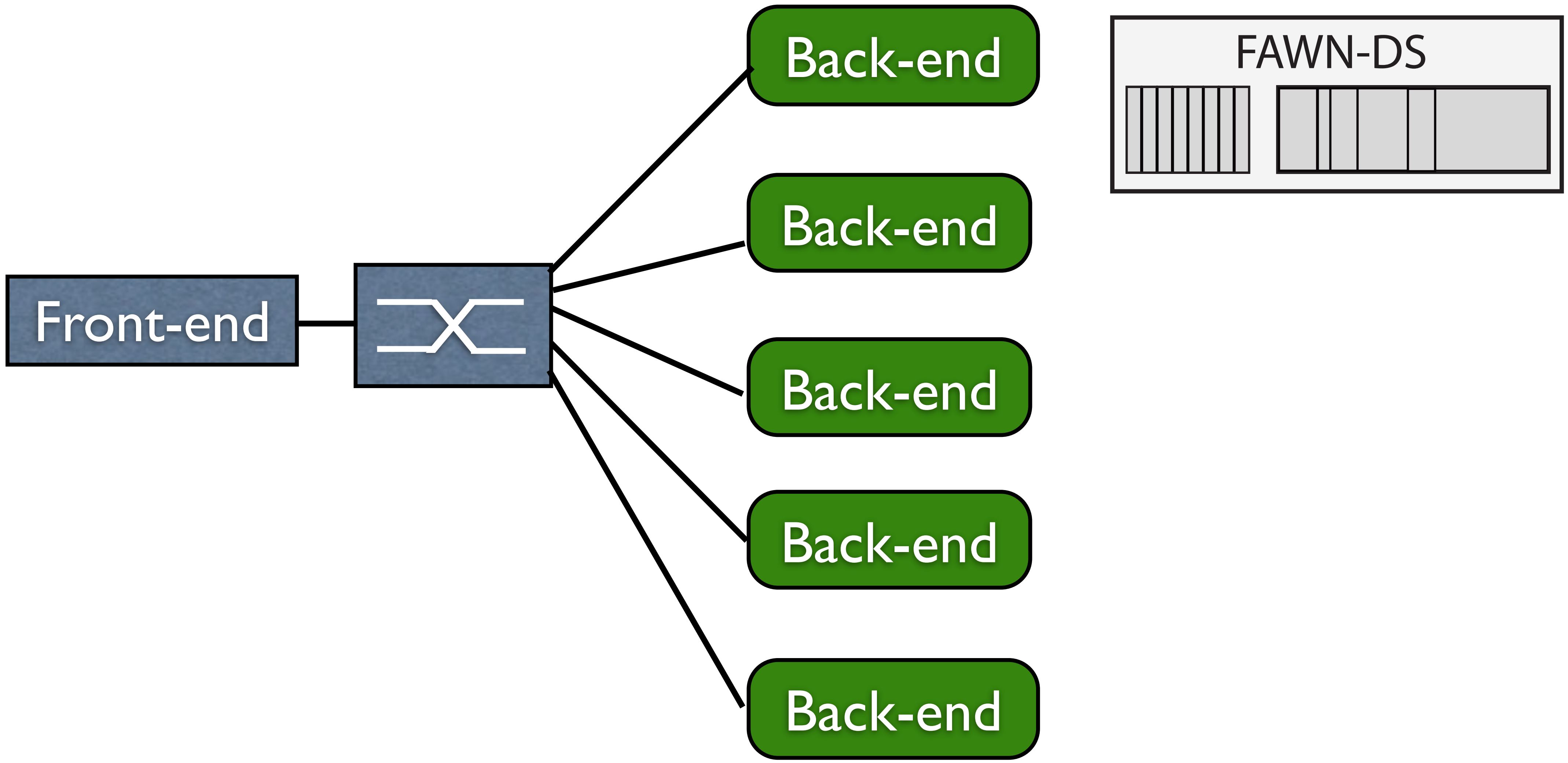






Consistent hashing

FAWN-KV Architecture



FAWN-DS

Limited Resources ☐
Avoid random writes ☐

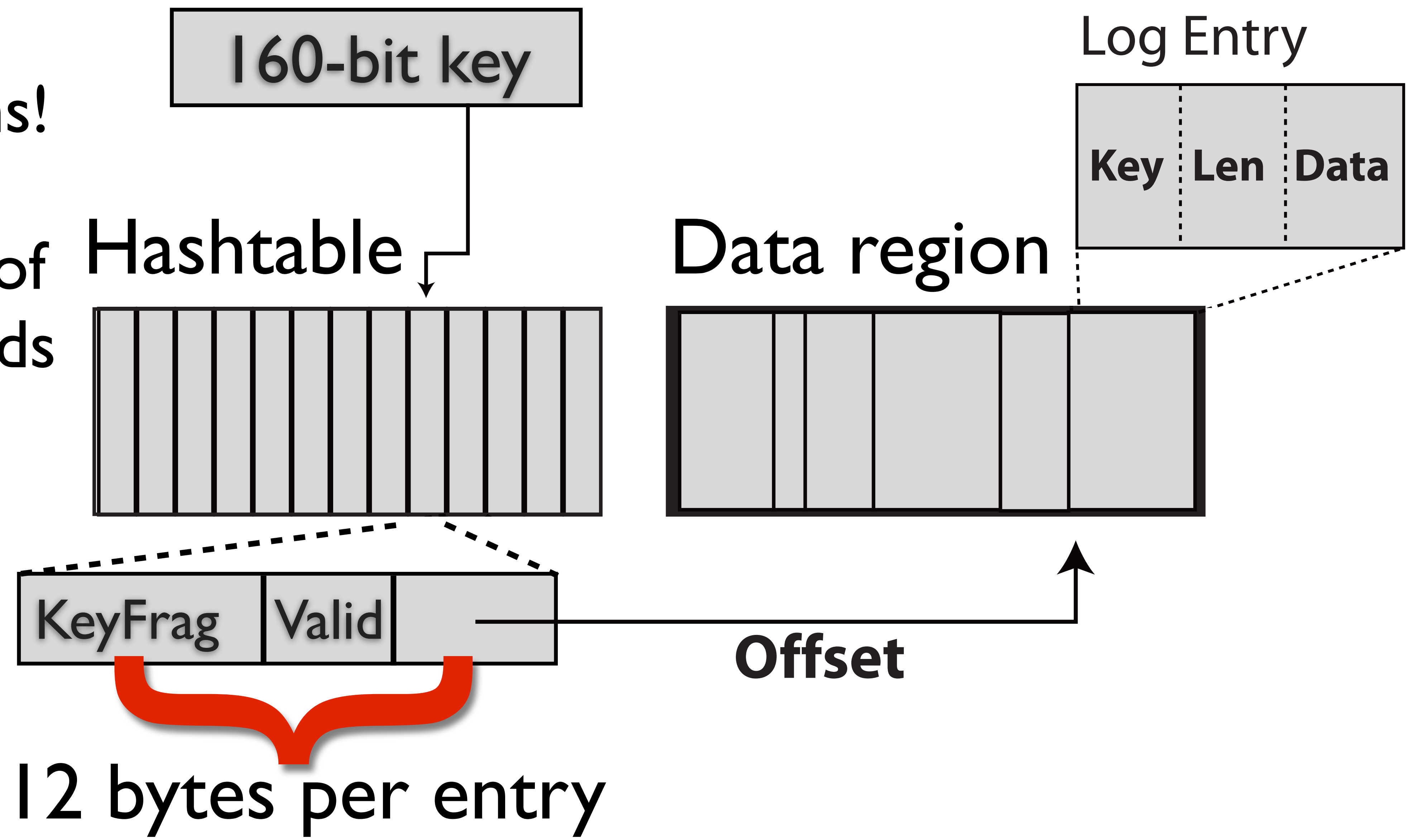
FAWN-KV

Efficient Failover ☐
Avoid random writes ☐

From key to value

KeyFrag != Key
Potential collisions!

Low probability of multiple Flash reads



FAWN-DS

- [x] Limited Resources
- [ ] Avoid random writes

FAWN-KV

- [ ] Efficient Failover
- [ ] Avoid random writes

Log-structured Datastore

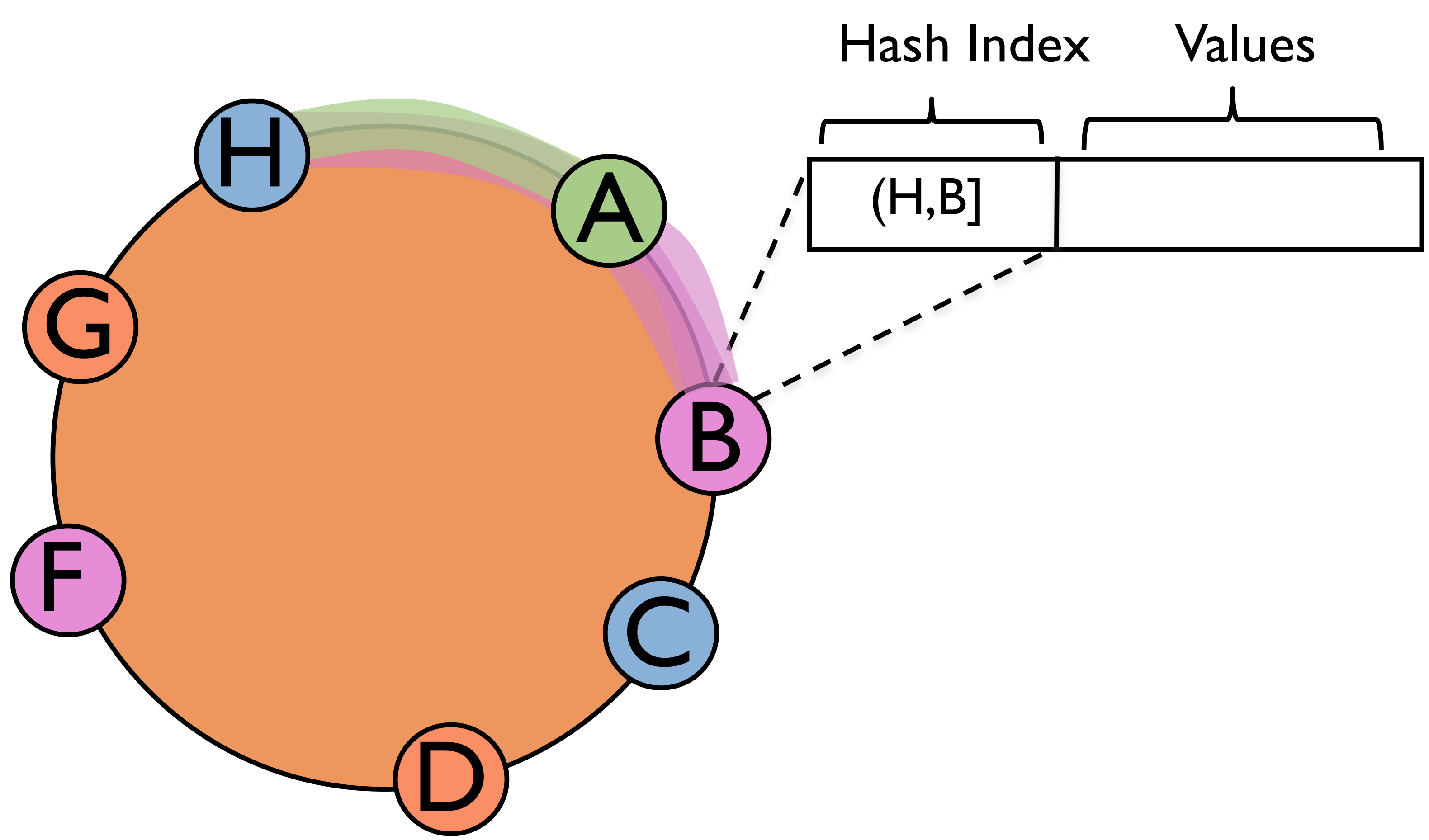
- Log-structuring avoids small random writes

Get → Random Read
Put → Append
Delete → Append

FAWN-DS	
Limited Resources	☑
Avoid random writes	☑

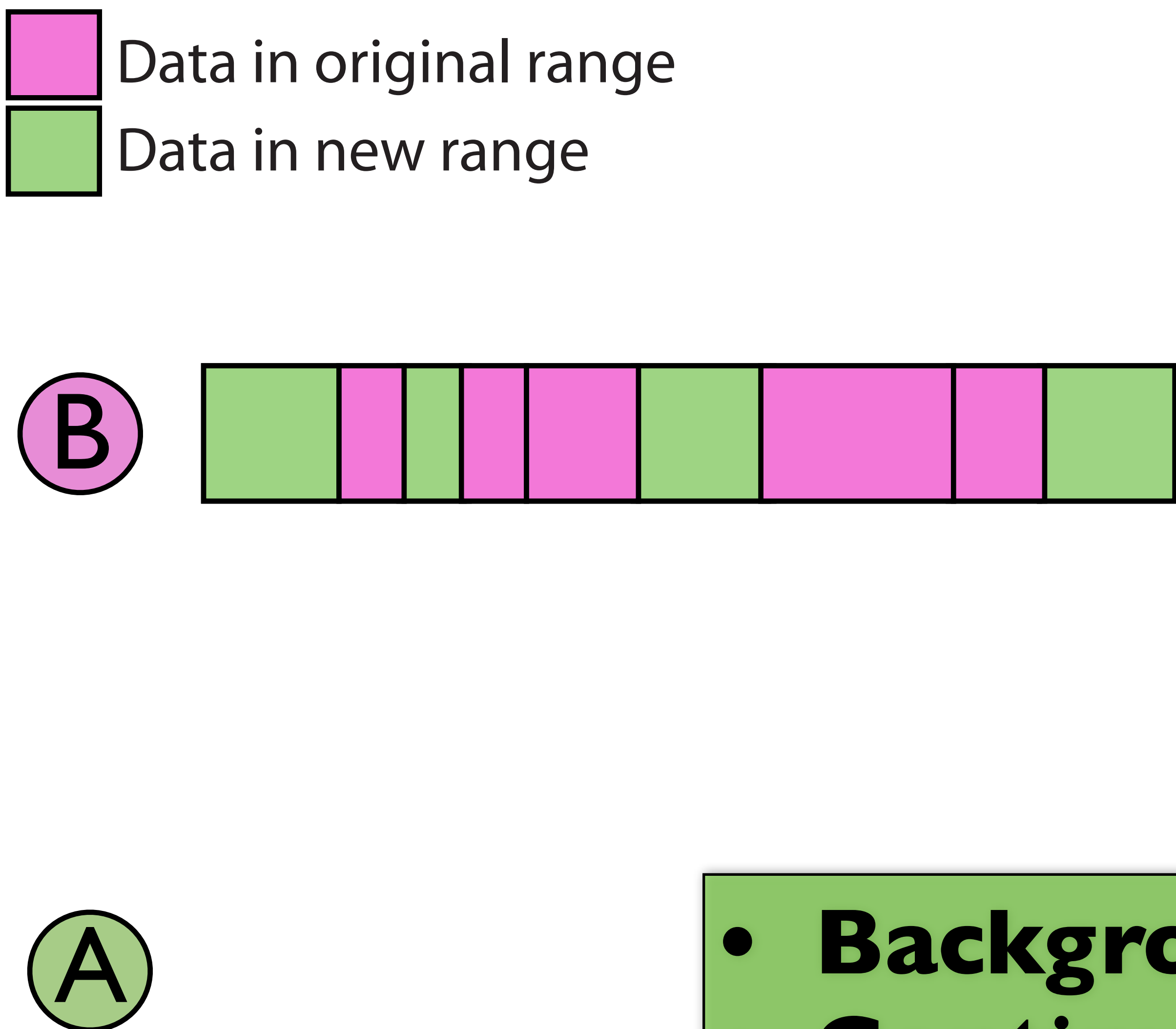
FAWN-KV	
Efficient Failover	☐
Avoid random writes	☐

On a node addition



Node additions, failures require transfer of key-ranges

Nodes stream data range



Stream from B to A

Concurrent Inserts,
Minimizes locking

Compact Datastore

- **Background operations sequential**
- **Continue to meet SLA**

FAWN-DS

Limited Resources ☑
Avoid random writes ☑

FAWN-KV

Efficient Failover ☑
Avoid random writes ☑

FAWN-KV Take-aways

- Log-structured datastore
  - Avoids random writes at all levels
  - Minimizes locking during failover
- Careful resource use but high performing
- Replication and strong consistency
  - Variant of chain replication (see paper)

Overview

- Background
- FAWN principles
- FAWN-KV Design
- **Evaluation**
- Conclusion

Evaluation Roadmap

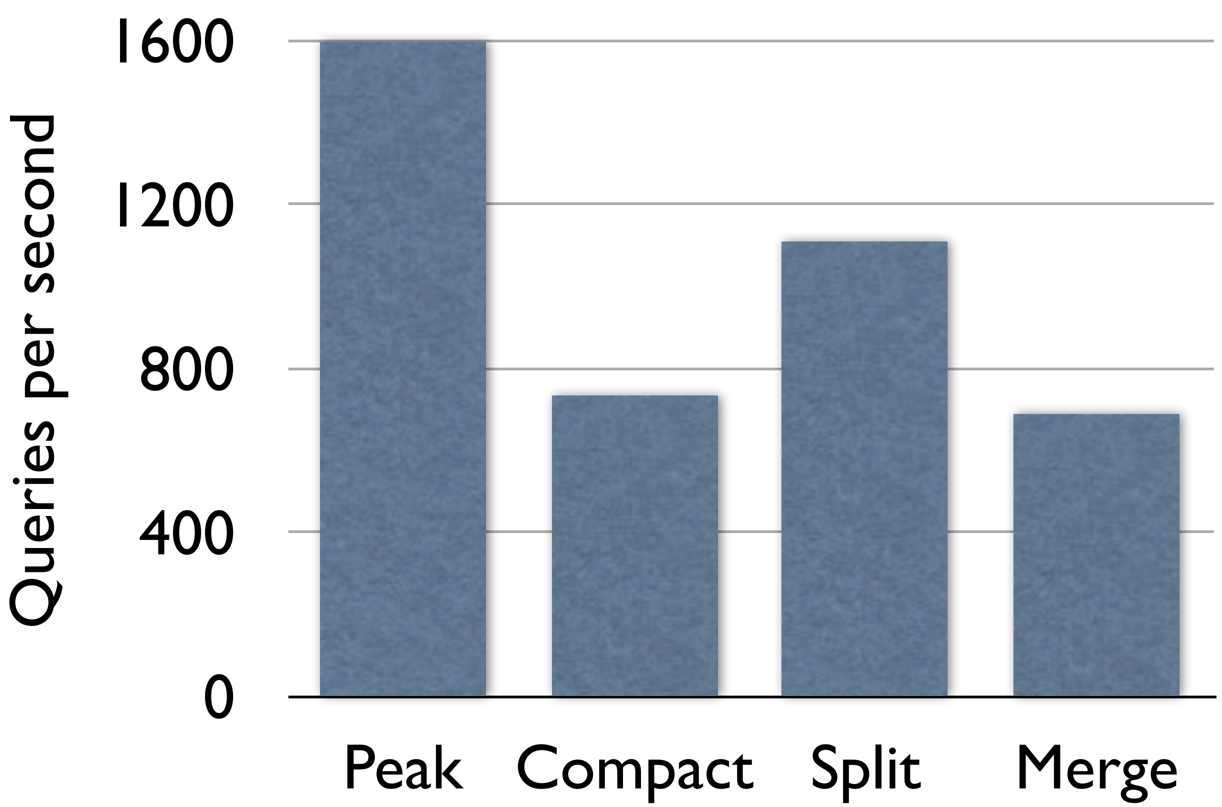
- Key-value lookup efficiency comparison
- Impact of background operations
- TCO analysis for random read workloads

FAWN-DS Lookups

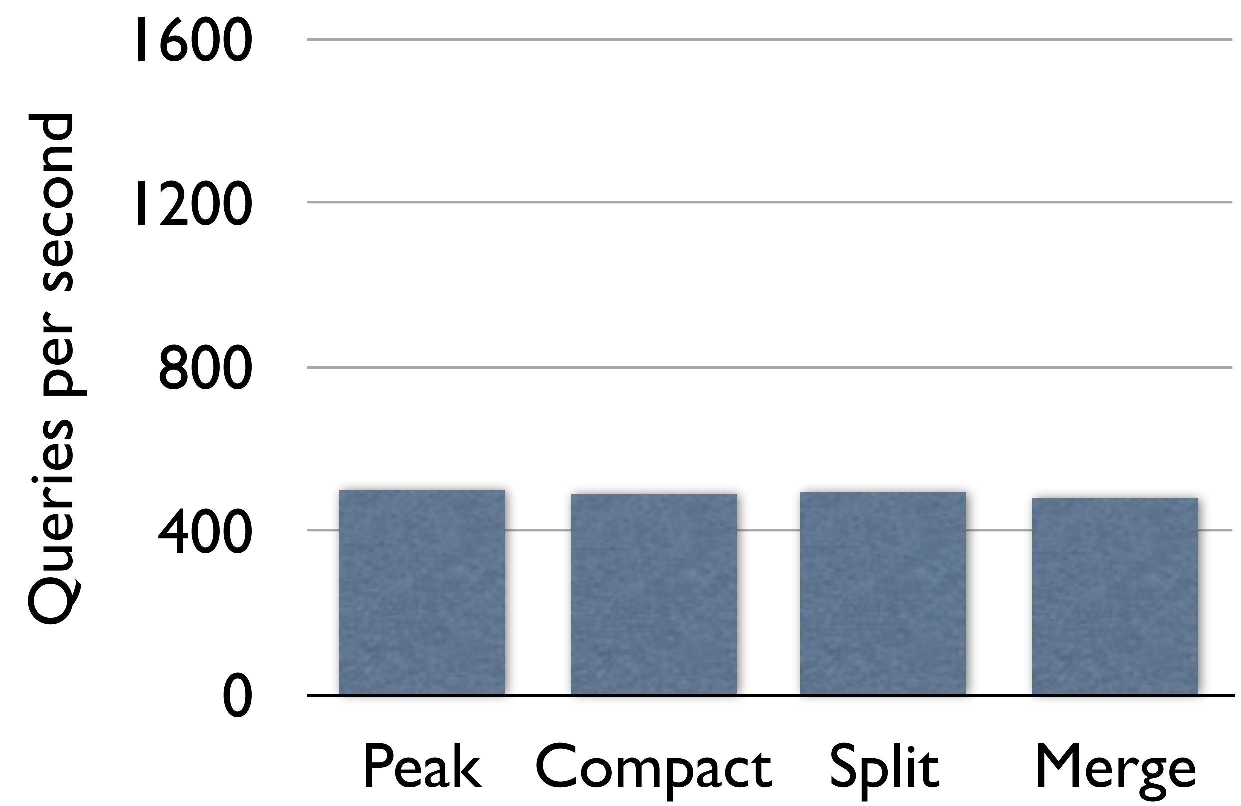
System	QPS	Watts	$\frac{QPS}{Watt}$
Alix3c2/Sandisk(CF)	1298	3.75	346
Desktop/Mobi (SSD)	4289	83	51.7
MacbookPro / HD	66	29	2.3
Desktop / HD	171	87	1.96

- Our FAWN-based system over 6x more efficient than 2008-era traditional systems

Impact of background ops



Peak query load



30% of peak query load

Background operations have:

- **Moderate** impact at peak load
- **Negligible** impact at 30% load

When to use FAWN for random access workloads?

TCO = Capital Cost + Power Cost ($0.10/kWh)

Traditional (200W)
Five 2 TB disks
160GB PCI-e Flash SSD
64GB FBDIMM per node

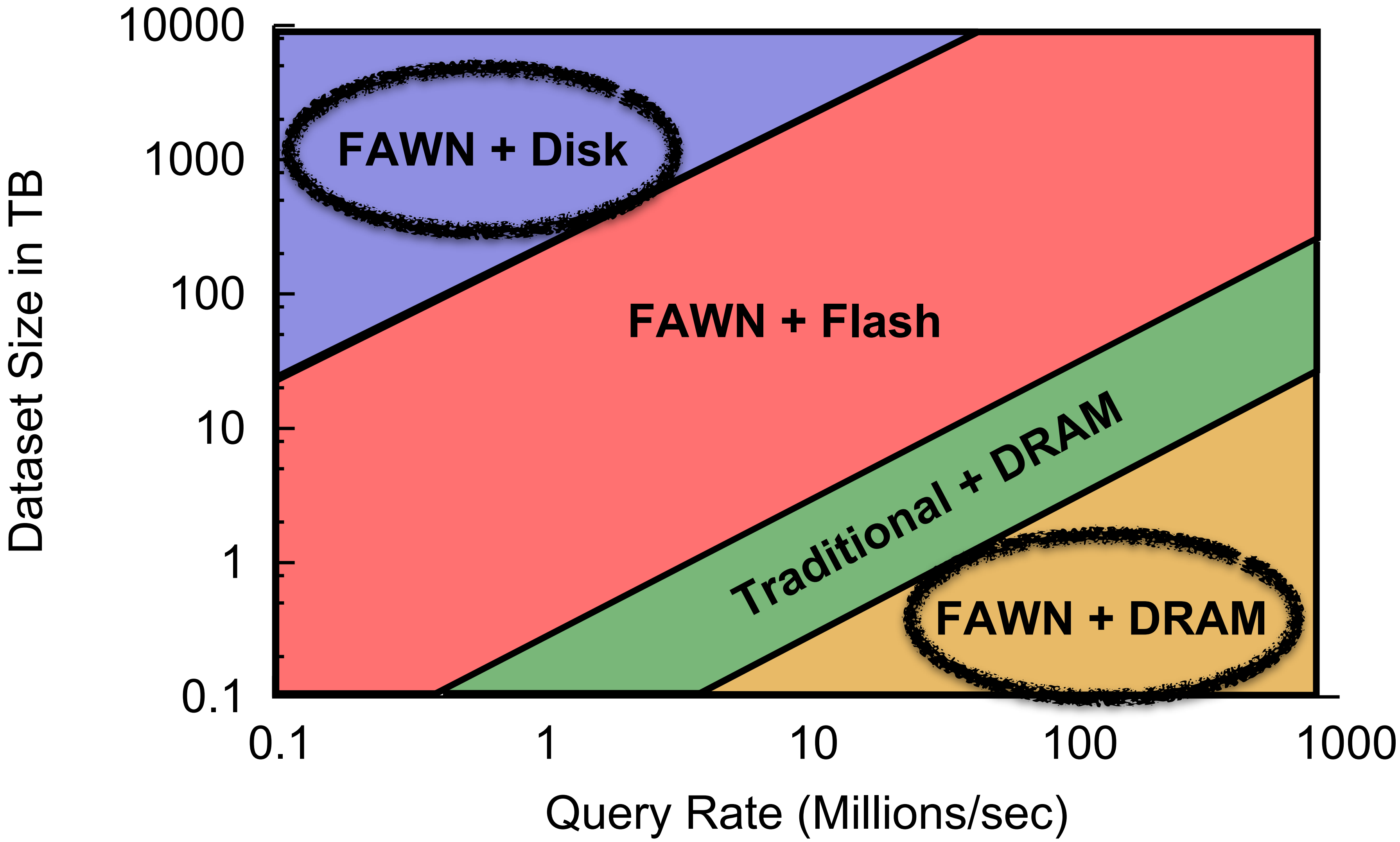
~$2000-8000 per node

FAWN (10W each)
2 TB disk
64GB SATA Flash SSD
2GB DRAM per node

~$250-500 per node

Architecture with lowest TCO

for random access workloads



Ratio of query rate to dataset size determines storage technology

Graph ignores management, cooling, networking...

FAWN-based systems can provide lower cost per {GB, QueryRate}

Conclusion

- FAWN architecture reduces energy consumption of cluster computing
- FAWN-KV addresses challenges of wimpy nodes for key value storage
  - Log-structured, memory efficient datastore
  - Efficient replication and failover
  - Meets energy efficiency and performance goals
- *"Each decimal order of magnitude increase in parallelism requires a major redesign and rewrite of parallel code" - Kathy Yelick*