

# Energy and Performance: Two Win-Win Examples

**Yitzhak Birk and Tomer Kol\***  
**Technion**  
**Haifa, Israel**

**[www.ee.technion.ac.il](http://www.ee.technion.ac.il)**

\* Tomer is presently with Google

# Overview

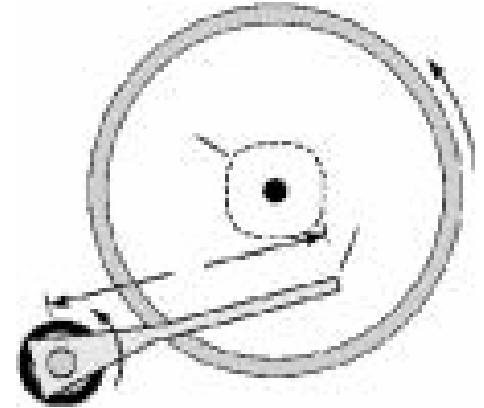
- **“Performance” and energy conservation are often conflicting goals.**
- **Key insight: reducing the amount of work (resources) required in order to achieve a given task is good for both.**
- **Our examples:**
  - **Web server:** “packaging” reduces the number of disk accesses for retrieving the contents of a web page.
  - **Transaction processing using satellite-based networks (+ sensor networks):** reducing the mean number of transmissions per transaction.

# Work Per Mission

- **Unit: seconds (time), but**
- **Unlike latency, this is actually (resource x time), like “person hours”**
- **If a resource consumes power, reducing the amount of work per given mission reduces energy consumption for a given workload.**

# Example 1: Web Server

- **Web server:**
  - Stores data on disk
  - Retrieves it in response to requests
- **Web page:**
  - HTML “skeleton”
  - Multiple small embedded objects
- **Each page requires multiple disk accesses:**
  - Wasted disk time → low maximum throughput
  - Extra disk seeks and time → more energy per workload
- **Same thing bad for both**  
⇒ same remedy good for both?




# PSL Home Page


[www.psl.technion.ac.il](http://www.psl.technion.ac.il)

- **Total size: ~150kB**
- **Composition:**
  - Skeleton
  - 8 embedded objects
- **Disk work:**
  - ~80ms
  - 9 seeks
- **Max disk throughput:**  
**~ 12 pages/sec**

About Us Page 1 of 2



**PARALLEL SYSTEMS LABORATORY**  
Department of Electrical Engineering

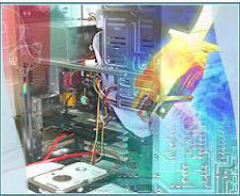


- Home
- **About Us**
- project proposals
- ongoing projects
- recent projects
- Past Projects
- Staff
- Sponsors & Donors

## ABOUT US

---

Established in  
October 1992



This Laboratory has as its primary goals research and education in the design and utilization of computer systems and subsystems in general, and parallel ones in particular

**Areas of activity presently include :**

- Components of operating systems.
- Storage-server architectures for Multi-media applications
- Performance evaluation of processors .
- Memory and storage and subsystems.
- Multi- and many-core architectures and applications.
- Architectures for new storage technologies.

**The equipment in the Laboratory includes:**

Several high-performance PCs (based on Pentium 4 processor)

Some multiprocessor systems for Infiniband based projects

High performance Symmetrix RAID System (500 GB) and Serial ATA Development Kits.

In addition,

we have various powerful simulation software packages,

# Possible Approaches

- **Prefetching:**
  - Mostly latency hiding, not a reduction in disk work per web page
  - (Smart scheduling can reduce mean seek distance and thus some reduction in work)
- **Contiguous placement of page's objects:**
  - Rotational latency not saved
  - Seek may not be served if server is busy, as other requests are interlaced with those of page
  - Not effective unless also read together

# Packaging

- **All page's objects:**
  - placed contiguously on disk
  - read in a single disk access.
- **After reading from disk:**
  - Server separates in memory (unaware client), or
  - sends entire package to client (participating client).
- **Price: replication of objects that are part of multiple pages (negligible or don't do it)**
- **Complication: need to update upon object change**
  - Server uses “check if modified since” after reading package
  - Various policies are possible upon change, ranging from “do nothing” to look for all copies and update them.

# Packaging: Impact (PSL example)

- 9 seeks → 1 seek per page
- Disk work: 80ms → 10ms
- Disk throughput: 12 pages/sec → 100 pages/sec
- Disk energy per web page:
  - Seek: down 9X
  - Motor: down ~9X (same power but for a shorter time)
  - Electronics: some savings

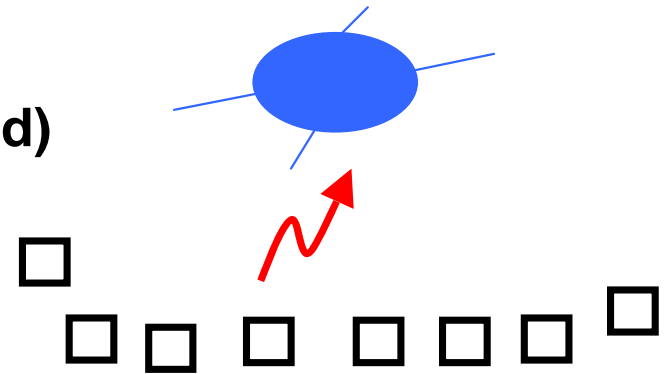
**Packaging increases throughput & saves energy!**



## Example 2: Maximizing Deadline-Constrained Capacity in Multi-channel ALOHA Networks

- **Multichannel ALOHA:**

- Upstream contention channels (shared)
- Private downstream channel for hub transmissions and Acks
- New message: draw random channel, transmit and wait for Ack
- If no Ack, redraw channel and retransmit



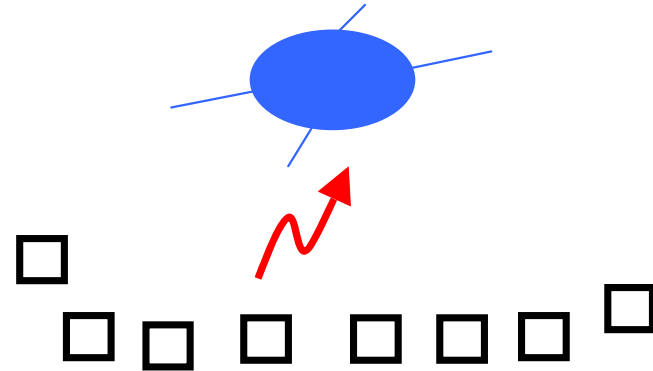
- **Transaction processing – rule of the game:**

- User: deadline and permissible  $\Pr(\text{failure})$
- Service provider's goal: max. transactions/sec

- **For battery-operated terminals and sensors: minimum mean energy per transaction.**

# Multi-Channel ALOHA: Example

- Scenario:
  - Delay permits 2 attempts (rounds).
  - $P(\text{collision}) = 0.1$
  - $P(\text{failure}) = 0.0001$
- Greedy approach: send 4 copies in 1st attempt:
  - minimum mean delay
  - 4 copies per message



- Better approach:
  - Send 1 in first attempt
  - Send 3 in 2<sup>nd</sup> attempt iff 1<sup>st</sup> fails
- Analysis
  - longer mean delay (who cares?)
  - 1.3 copies per message
  - 3-fold higher capacity
  - 3x less energy per transaction!

Higher Throughput and Less Energy!

# Conclusions

- **Careful attention to the true performance goals can be extremely beneficial**
- **While energy reduction may be at odds with other goals, reduction of the amount of work per mission can create win-win situations**
- **Our examples combined these observations!**