Energy-Efficiency of Long-term Storage

Irina Tolokonnikova

Seminar "Energy-Efficient Programming"
Arbeitsbereich Wissenschaftliches Rechnen
Fachbereich Informatik
Fakultät für Mathematik, Informatik und Naturwissenschaften
Universität Hamburg

2015-01-14

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

informatik
die zukunft

2015-03-01

Energy-Efficiency of Long-term Storage

Energy-Efficiency of Long-term Storage

Irina Tolokonnikova

Seminar "Energy-Efficient Programming"
Arbeitsbereich Wissenschaftliches Rechnen
Fachbereich Informatik
Fakultät für Mathematik, Informatik und Naturwissenschaften
Universität Hamburg

2015-01-14

Agenda

Agenda
1 Archive
2 Data Storage Devices
data storage methods
tape
HDD
MAIDs
3 State of Research
4 Conclusion
5 References

show the audience what and how much data an archive has to handle with. I want to use the data of the DKRZ given on it's homepage and the xkcd-what if? about google
https://what-if.xkcd.com/63/

What is an archive for digital data? How is data stored? and about how much data are we talking about?

Archive

- storage of digital data for many years
- requirements:
  - preservation
  - retrieval
  - auditing
- archival data $\neq$ backup data
- needs to be cheap to obtain, cheap to operate, easy to expand
- high costs for energy consumption
  $\rightarrow$ room for improvement

Archive
storage of digital data for many years
requirements:
preservation
retrieval
auditing
archival data $\neq$ backup data
needs to be cheap to obtain, cheap to operate, easy to expand
high costs for energy consumption
$\rightarrow$ room for improvement

main requirements of an archivebut the main problem of archives are the high costs for energy consumption, and in this talk, we want are going to see, how we can reduce this costsimprovements are important in this field, because we want to match the requirements for big data

that means, we want to work with huge amounts of data, compare it and search for association rules e.g.

Google

- How much data are we talking about?

Google
How much data are we talking about?

how much data is stored in such an archive?first the data from the DKRZ homepage, to have an anchor for further comparissonsestimashion based on the published energy consumption. Google doesn't publish how much data they store.
what if-comic, where peaple can send in absurde but interesting questions, and Randall Munroe will try to answer it. that's more than a lot of data, because in 2013 only 8 ExaBytes of Hard Drives were produced for sale in total worldwide

NSA probably stores 1 YotaByte

Google

- How much data are we talking about?
  - DKRZ: > 100 PetaBytes total capacity [1]

Google
How much data are we talking about?
DKRZ: > 100 PetaBytes total capacity [1]

how much data is stored in such an archive?first the data from the DKRZ homepage, to have an anchor for further comparissonsestimashion based on the published energy consumption. Google doesn't publish how much data they store.
what if-comic, where peaple can send in absurde but interesting questions, and Randall Munroe will try to answer it. that's more than a lot of data, because in 2013 only 8 ExaBytes of Hard Drives were produced for sale in total worldwide

NSA probably stores 1 YotaByte

Google

- How much data are we talking about?
  - DKRZ: $>$ 100 PetaBytes total capacity [1]
  - Google: $\sim$ 15 ExaBytes (in 2013) = 15000 Petabytes (only estimation)

Google
- How much data are we talking about?
  - DKRZ: > 100 PetaBytes total capacity [1]
  - Google: ~ 15 ExaBytes (in 2013) = 15000 Petabytes (only estimation)

how much data is stored in such an archive?first the data from the DKRZ homepage, to have an anchor for further comparissonsestimashion based on the published energy consumption. Google doesn't publish how much data they store.
what if-comic, where peaple can send in absurde but interesting questions, and Randall Munroe will try to answer it. that's more than a lot of data, because in 2013 only 8 ExaBytes of Hard Drives were produced for sale in total worldwide

NSA probably stores 1 YotaByte

Google

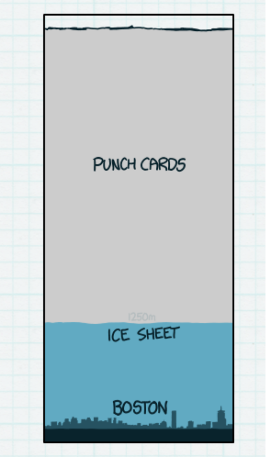


Figure: 15 ExaBytes of punch cards would be enough to cover New England, to a depth of about 4.5 kilometers

Google

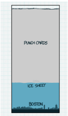


Figure: 15 ExaBytes of punch cards would be enough to cover New England, to a depth of about 4.5 kilometers

but which device should we coose, if we want to store as much data as google does?

show some devices, that are not uses for archives and which are

Show a real cassette, make clear, why LP or Punch cards are not suitible for long term storage

not this

Figure: LP [wikipedia.org]

probably not LPs - not easy to search and not as much data, BUT it lasts very long without errors, because it is engraved

LPs were send to space, so alien life forms get an impression of the earth, music and human life. It was well chosen, because it had to have a long lifetime

not this

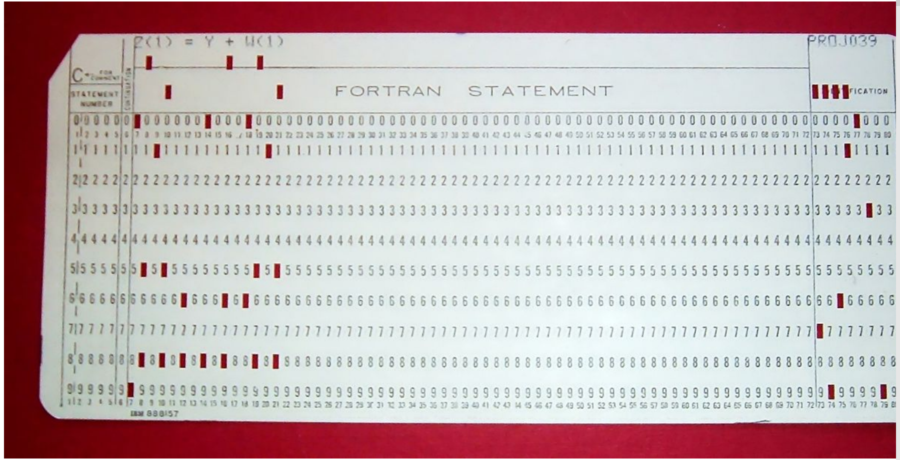


Figure: punch card [wikipedia.org]

not this

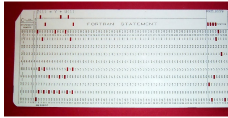


Figure: punch card [wikipedia.org]

as we allready learned, punch cards would not be suitible for masses of data

not this

Figure: a United States National Archives Records Service facility in 1959. Each carton could hold 2000 cards [wikipedia.org]

not this

Figure: a United States National Archives Records Service facility in 1959. Each carton could hold 2000 cards [wikipedia.org]

… because than out storage would look like this

not this

Figure: 3,5-inch floppy disk

Show, how "normal" cassettes look like and where we know them in every days life. maybe you still know cassetts, but do your younger siblings still know how to use them? explain, that they are still used in another region

not this?

Figure: compact cassette [wikipedia.org]

not this?

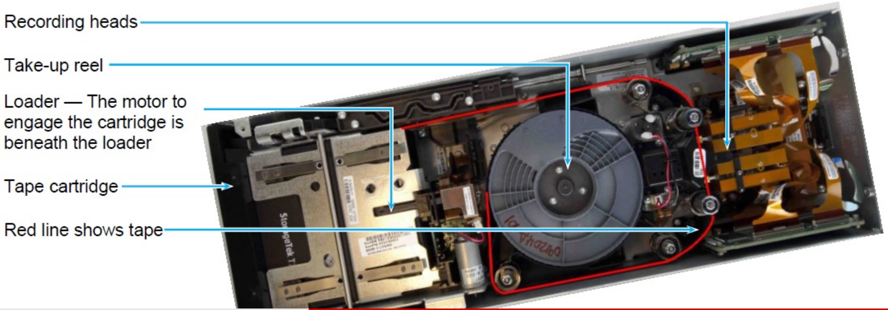
Figure: compact cassette [wikipedia.org]

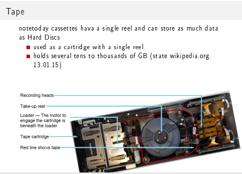
maybe some of you can remember, that cassetts were used not only for music but also for data in such computers like C64 but the casstetts used for data storage have changed

Tape

notetoday cassettes hava a single reel and can store as much data as Hard Discs

- used as a cartridge with a single reel
- holds several tens to thousands of GB (state wikipedia.org 13.01.15)



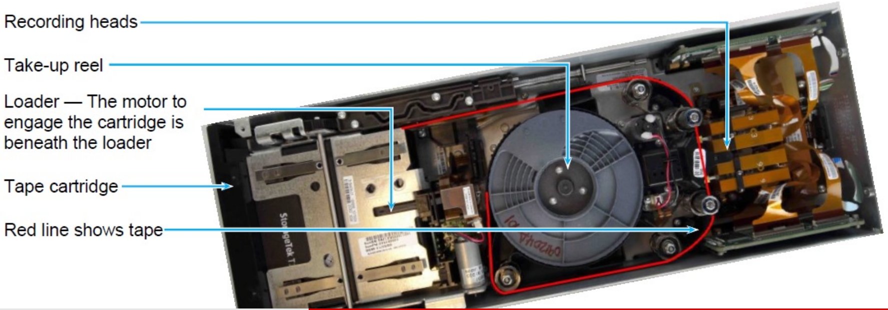


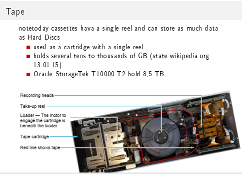
they look like this and are well protected by the case

Tape

notetoday cassettes hava a single reel and can store as much data as Hard Discs

- used as a cartridge with a single reel
- holds several tens to thousands of GB (state wikipedia.org 13.01.15)
- Oracle StorageTek T10000 T2 hold 8,5 TB





they look like this and are well protected by the case

DKRZ

- 7 automated Oracle/StorageTek SL8500 tape libraries
- 8 robots per library
- over 67000 slots for magnetic tape cassettes

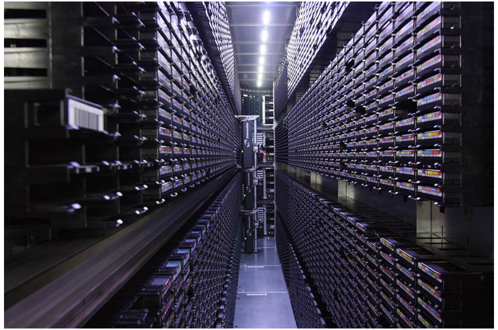


Figure: Inside the Tape library of DKRZ [1]

DKRZ
- 7 automated Oracle/StorageTek SL8500 tape libraries
- 8 robots per library
- over 67000 slots for magnetic tape cassettes

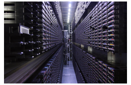


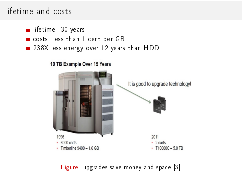
Figure: Inside the Tape library of DKRZ [1]

main summery about tape

lifetime and costs

- lifetime: 30 years
- costs: less than 1 cent per GB
- 238X less energy over 12 years than HDD

Figure: upgrades save money and space [3]



main point, why tape is used for long term storageWhen not in use, tape doesn't produce any energy and doesn't need any ether it is important to upgrade the device, even if it hase a long lifetime, because you can save much space, as shoven in this graphic

pros and cons

Pros	Cons
cheap	needs special expensive equipment
long lifetime	sequential access pattern
no power needed when not accessed	

pros and cons

Pros	Cons
cheap	needs special expensive equipment
long lifetime	sequential access pattern
no power needed when not accessed	

Tape has to be rewinded when searched and needs much more time to get to information, even if you know, where it is stored.

Even more time is needed, if you want to search for data or compare it.

Hard Drives

- easy and fast to access data storage
- searching, consistency checking and inter-media reliability operations
- costs: 0.07 $per GB and falling
- lifetime: 10 years, but easy to break mechanics

Figure: Laptop Hard disk drive, 500 GB Western Digital Scorpio

Same for Hard drives, but in this case a real object is not needed, because every one of you probably has one ore more at homethe costs get higher, when we need to change the Discs more often and when we need redundancy to save data, in case one discs breakesnormal HDD, like in most of your laptops probably

pros and cons

Pros	Cons
easy access, simply system	needs much power, even when turned off
matches requirements of big data	easy to break
higher bandwidth (200X)	needs extra space for redundancy

pros and cons

Pros	Cons
easy access, simply system	needs much power, even when turned off
matches requirements of big data	easy to break
higher bandwidth (200X)	needs extra space for redundancy

summery about HDD, main points relevant for long term storage when storing on HDD the main point to remember is, that extra space for redundancy is need, and that's why more space is needed for the devices, than we would estimate from raw data amount.

Colarelli, Grunwald et al.(2002)

- massive array of idle disks = MAIDs
- aim: storage densities matching those of tape, with reduced energy consumption
- but operating same data volume in disks costs 10X more than in tape
- idea: use a cache manager to keep only part of disks in an array powered up
- varying spin-down delays

Colarelli, Grunwald et al.(2002)
- massive array of idle disks = MAIDs
- aim: storage densities matching those of tape, with reduced energy consumption
- but operating same data volume in disks costs 10X more than in tape
- idea: use a cache manager to keep only part of disks in an array powered up
- varying spin-down delays

a newer approach on handling HDD energy efficiently. an important article, cited in following research articles
dealing with the question, how we can store on HDD more efficiantly, maybe someday as eddiciant as on tapewhen using only the disks already spinning, we can save energy.

previos observation: mostly searched information is thoes just saved or added. So by keeping this information in cache we probably can manage most of the requests ...to save extra energy

Results

- good trade off in performance and energy efficiency
- read performance still effected by the spin-down delay
- but 82% of read requests were satisfied by the cache
- least energy consumed with 4 sec spin-down delay

Results

- good trade off in performance and energy efficiency
- read performance still effected by the spin-down delay
- but 82% of read requests were satisfied by the cache
- least energy consumed with 4 sec spin-down delay

what are the main points about MAIDsusing MAIDs the HDD might still not be as efficiant as tape, but it's a good approach to work on

SSD

- costs: 0.66 $per GB , yet too expensive
- lifetime depends on usage, ∼10 years
- yet unclear, how unused data behaves on SSD
- coming soon?

SSD

- costs: 0.66 $per GB , yet too expensive
- lifetime depends on usage, ∼10 years
- yet unclear, how unused data behaves on SSD
- coming soon?

looking forward: will we use SSD in some years?
other possible storage devices are coming, but that it is not yet clear, if good for storage and how fast information gets lost on SSD

Pergamum tomes by Storer et al. (2008)

- interfaces and protocols change slowly
- using inter- and intra-device redundancy
- work energy efficient, by not spinning up idle disks

→ intelligent, self managing storage device

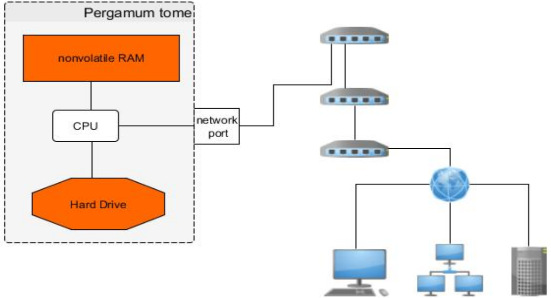
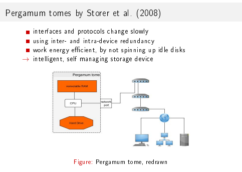


Figure: Pergamum tome, redrawn



newer research and what will maybe be the future of long term storage
talk about, how HDDs can be made more efficient but still fit Big Data
Named after the library of Alexandria

main ideas of this paper are based on thoes about MAIDsUse the fact, that interfaces and protocols take much more time to change. Just thin of http ... not only for restoring data, which was saved on broken devices, but also for reading

Results of Pergamum

- size of the hard drive
- nonvolatile RAM handles many types of requests(e.g. hashes) without spinning up the disk
- using signatures for redundancy checking in entire inter-disk group
- using trees of hash values to reduce signature data
- once added to the network, the tome automatically joins a redundancy group or builds new one

→ makes storage management easier

- using intra-device redundancy, recovering from small errors without other devices
- aim to be price-competitive with tape

Results of Pergamum
- size of the hard drive
- nonvolatile RAM handles many types of requests(e.g. hashes) without spinning up the disk
- using signatures for redundancy checking in entire inter-disk group
- using trees of hash values to reduce signature data
- once added to the network, the tome automatically joins a redundancy group or builds new one
→ makes storage management easier
- using intra-device redundancy, recovering from small errors without other devices
- aim to be price-competitive with tape

main points about the Pergamum tome, that can help improve it in futuresmall and cheapenergy efficiant itselfgood algorithm helps reducing the energyonly a minimum trained administrator needed to change the broken devices once a month.

the new devices will organise themselves in existing groups or start new onesthis goal is not reached (yet?)

Problems and improvments

- still not included in data archives(?)
- redundancy overhead, but much energy saved
- "disposable" tomes
- encoding time 10X longer than on laptop processor BUT 10X less power consumed
- future work:
  - better algorithms
  - parallel processes (distributed searching)

Problems and improvments

- still not included in data archives(?)
- redundancy overhead, but much energy saved
- "disposable" tomes
- encoding time 10X longer than on laptop processor BUT 10X less power consumed
- future work:
  - better algorithms
  - parallel processes (distributed searching)

I don't know much about this point, and I couldn't find anything about thisare desposable devices realy more energy efficiant and saving money? Do we realy want so throw that much HHDs away?future tasks named in paper

A Spin-Up Saved is Energy Earned, Greenan et al.(2008)

- idea: use redundancies on active devices instead of waking up inactive ones

→ Power aware coding

- three conditions needed:

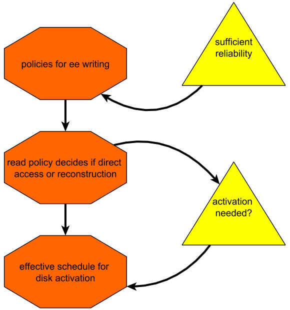
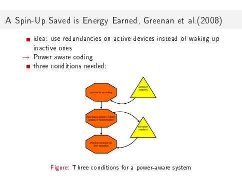


Figure: Three conditions for a power-aware system



based on "A spin-up saved is Energy earned" paper by Greenan et al. (2008). It's slightly newer and i also presents some algorithms. But I find it a little confusing and I think, it would be too much, if I explained it in detail, so here is only a summary of the main ideas, with pictures I made to illustrate them

Power Aware Techniques

- rules known from Pergamum tome
- Power Schedule
  - each code instance should have own write policy
  - write parallel across disk groups
- Power-Aware Read Algorithm
  - minimize the number of disk activations
  - first find out, if lost data is recoverable
  - like solving a matrix where inactive devices are treated as erased
- Disk Activation Algorithm
  - perform search to find best activation
  - how and when is a spin-down performed?

Power Aware Techniques
- rules known from Pergamum tome
- Power Schedule
  - each code instance should have own write policy
  - write parallel across disk groups
- Power-Aware Read Algorithm
  - minimize the number of disk activations
  - first find out, if lost data is recoverable
  - like solving a matrix where inactive devices are treated as erased
- Disk Activation Algorithm
  - perform search to find best activation
  - how and when is a spin-down performed?

Power aware Techniques as named in paperalready used in Pergamum and learned from Storer et al., since this paper is also citing the Pergamum-paper

observation while testing

- mind the trade-off trilemma!

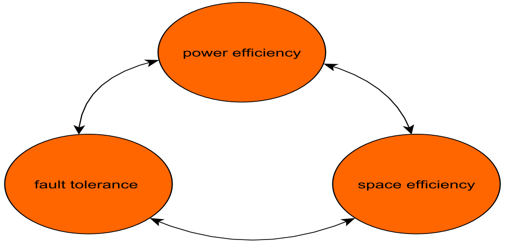
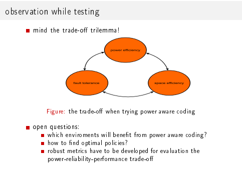


Figure: the trade-off when trying power aware coding

- open questions:
  - which enviroments will benefit from power aware coding?
  - how to find optimal policies?
  - robust metrics have to be developed for evaluation the power-reliability-performance trade-off



it's not called trilemma in the paper, that's my interpretation of it and the main point I want the audience to keep in mind:

You allways habe trade-offs, be aware where you want to have them.important point for future research

Conclusion

	Disk	Tape
Max shelf life (bit rot)	10 years	30 years
Best practices for data migration to new technology	3-5 years	8-12 years
Uncorrected Bit Error Rate, Probability (avg 1 error in x TB)	10^{-14} (~10's of TB)	10^{-19} (~1 million TB)
Power and cooling	238X	X

Figure: Disk compared to Tape [3]

Conclusion

	Disk	Tape
Max shelf life (bit rot)	10 years	30 years
Best practices for data migration to new technology	3-5 years	8-12 years
Uncorrected Bit Error Rate, Probability (avg 1 error in x TB)	10^{-14} (~10's of TB)	10^{-19} (~1 million TB)
Power and cooling	238X	X

Figure: Disk compared to Tape [3]

Conclusion

- Pergamum tomes by Storer et al.
  - Pergamum tomes added to networks
  - redundancy overhead used to recover errors
  - energy saved by not spinning up other disks
  - self managing system with "disposable" nodes
- Power Aware Programming
  - try to use less disks as efficient as you can
  - mind the trade-off trilemma between fault tolerance, space efficiency and power efficiency
  - *"Initial results show that power-aware coding may be well suited for the write-once, read-maybe workload of long-term archival storage systems."*

Conclusion
- Pergamum tomes by Storer et al.
  - Pergamum tomes added to networks
  - redundancy overhead used to recover errors
  - energy saved by not spinning up other disks
  - self managing system with "disposable" nodes
- Power Aware Programming
  - try to use less disks as efficient as you can
  - mind the trade-off trilemma between fault tolerance, space efficiency and power efficiency
  - *"Initial results show that power-aware coding may be well suited for the write-once, read-maybe workload of long-term archival storage systems."*

In the summery I want to present the pros and cons of tape and HDD and on which data or in which fildes which seems to be the better choise. Also I want to remind, that the life time of the device chosen should not be forgotten

this would be the basic slid for discussion this quote shows, that you always have to keep in mind, how and how often you want to reuse the stored information

if you want to rearch for data or compare random files or look for association rules e.g., you will have to calculate with higher energy consumption

How would you store...

- ...(your own) private medical data?
- ...research data of a medical study?
- ...data of all patients of a hospital?

How would you store...
- ...(your own) private medical data?
- ...research data of a medical study?
- ...data of all patients of a hospital?

Instead of a slide whit "Any questions? Thanks for listening" I want to end with a question, the audience should answer for themselves, which device they would choose.

How would you store...

- ...(your own) private medical data?
- ...research data of a medical study?
- ...data of all patients of a hospital?

Instead of a slide whit "Any questions? Thanks for listening" I want to end with a question, the audience should answer for themselves, which device they would choose.

References

[1] https://www.dkrz.de/Klimarechner-en/datenarchiv (13.01.2015)
[2] https://what-if.xkcd.com/63/ (13.01.2015)
[3] Dr. Mark L Watson: *Advanced Tape Technologies for Future Archive Storage Systems*. MSST - Media II (Tape Media and Libraries), 2013
[4] Colarelli, Dennis, Dirk Grunwald, and Michael Neufeld. *The case for massive arrays of idle disks (maid)*. The 2002 Conference on File and Storage Technologies. 2002.
[5] Storer, Mark W., et al. *Pergamum: Replacing tape with energy efficient, reliable, disk-based archival storage.* Proceedings of the 6th USENIX Conference on File and Storage Technologies. USENIX Association, 2008.
[6] Greenan, Kevin M., et al. *A Spin-Up Saved Is Energy Earned: Achieving Power-Efficient, Erasure-Coded Storage.* HotDep. 2008.

References
[1] https://www.dkrz.de/Klimarechner-en/datenarchiv (13.01.2015)
[2] https://what-if.xkcd.com/63/ (13.01.2015)
[3] Dr. Mark L Watson: Advanced Tape Technologies for Future Archive Storage Systems. MSST - Media II (Tape Media and Libraries), 2013
[4] Colarelli, Dennis, Dirk Grunwald, and Michael Neufeld. The case for massive arrays of idle disks (maid). The 2002 Conference on File and Storage Technologies. 2002.
[5] Storer, Mark W., et al. Pergamum: Replacing tape with energy efficient, reliable, disk-based archival storage. Proceedings of the 6th USENIX Conference on File and Storage Technologies. USENIX Association, 2008.
[6] Greenan, Kevin M., et al. A Spin-Up Saved Is Energy Earned: Achieving Power-Efficient, Erasure-Coded Storage. HotDep. 2008.