DeepDive & Analytics Systems

Christopher Ré

Course Announcements

- OHs: Note that we have / are shifting these to provide better coverage before Project Part 3 & the final!
  - See the calendar on the website
  - Come to OHs early- don't leave questions / problems till the last minute!

- Piazza @1253, "Final Exam Review Session- Topic requests"
  - Please suggest / vote on topics for the final review session!!!

- Course evaluations online- please fill these out ☺
  - We want your feedback to make the course even better next time!
  - I read **every** evaluation.

Course Announcements

- We received some feedback that examples used in the course might feel exclusionary to groups in the class
  - Not our intention!
  - Please keep giving us this kind of feedback.
  - We try very hard to consider these issues, and pick examples for lecture & assignments that are not more suited to some groups than others… but we are not always perfect
  - We’ll keep trying.

Research Lecture

1. A taste of new applications driving research ***(Dark Data)***

2. A crash course in data analytics ***(The Algorithm for Big Data?)***

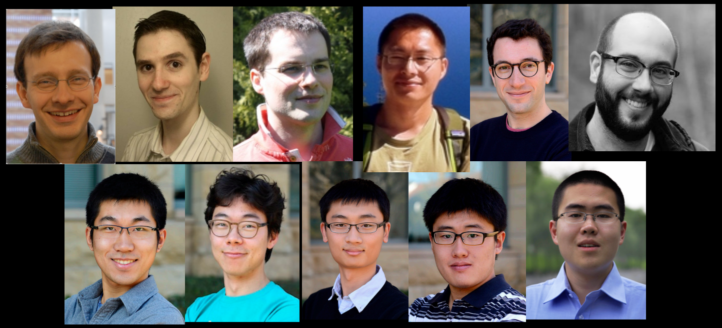
3. Hardware trends driving analytics systems

SHAMELESSLY BIASED BY MY GROUP'S WORK.

What you will learn about in this section

1. Dark Data Systems & DeepDive
2. ACTIVITY: Ask me questions!

DeepDive



The DeepDive Team

http://deepdive.stanford.edu/

Dark Data System: ETL on Steroids

Quality that can exceed paid human annotators and volunteers

DeepDive

Extraction, Integration, & Cleaning are ***inference problems***

Focus on key step for end-to-end application quality

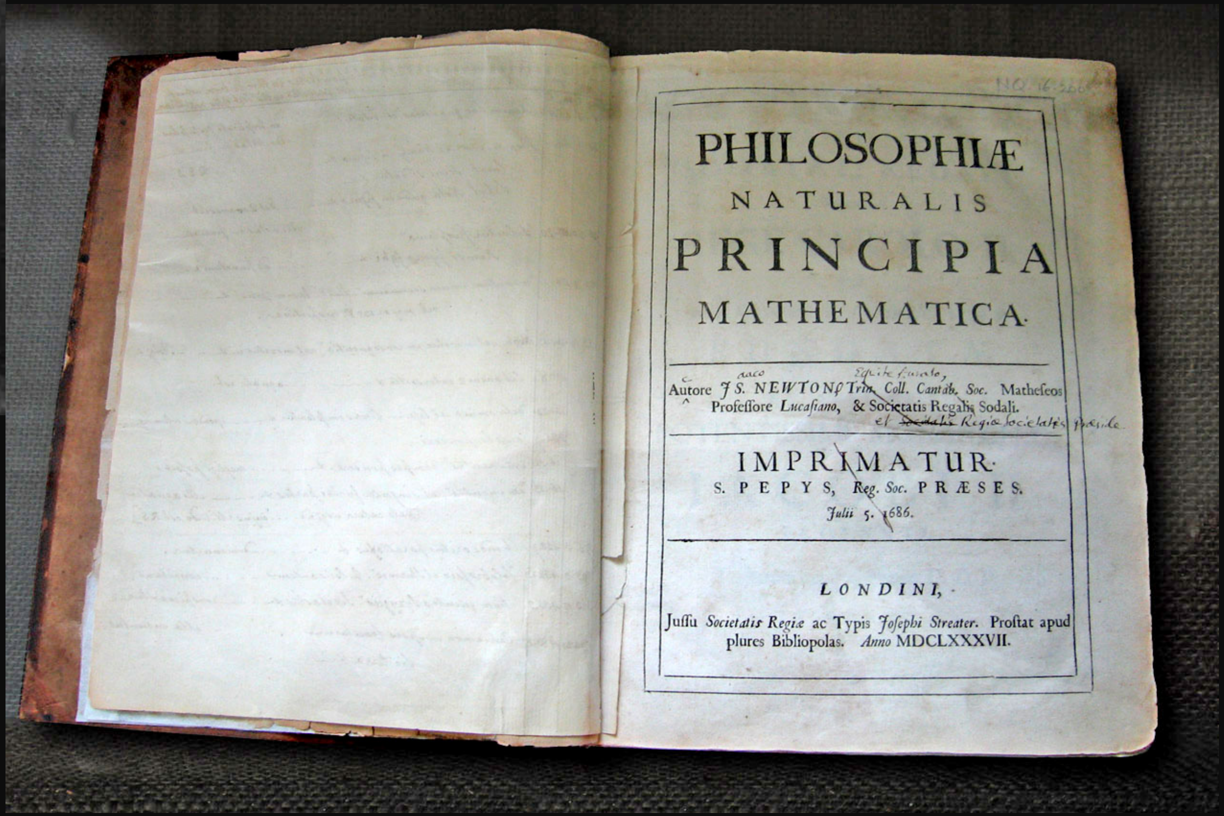
DeepDive

Enables non-CS users, but
scale is a challenge.

Modern hardware is changing how to
build analytics systems.

First, a story…

Newton

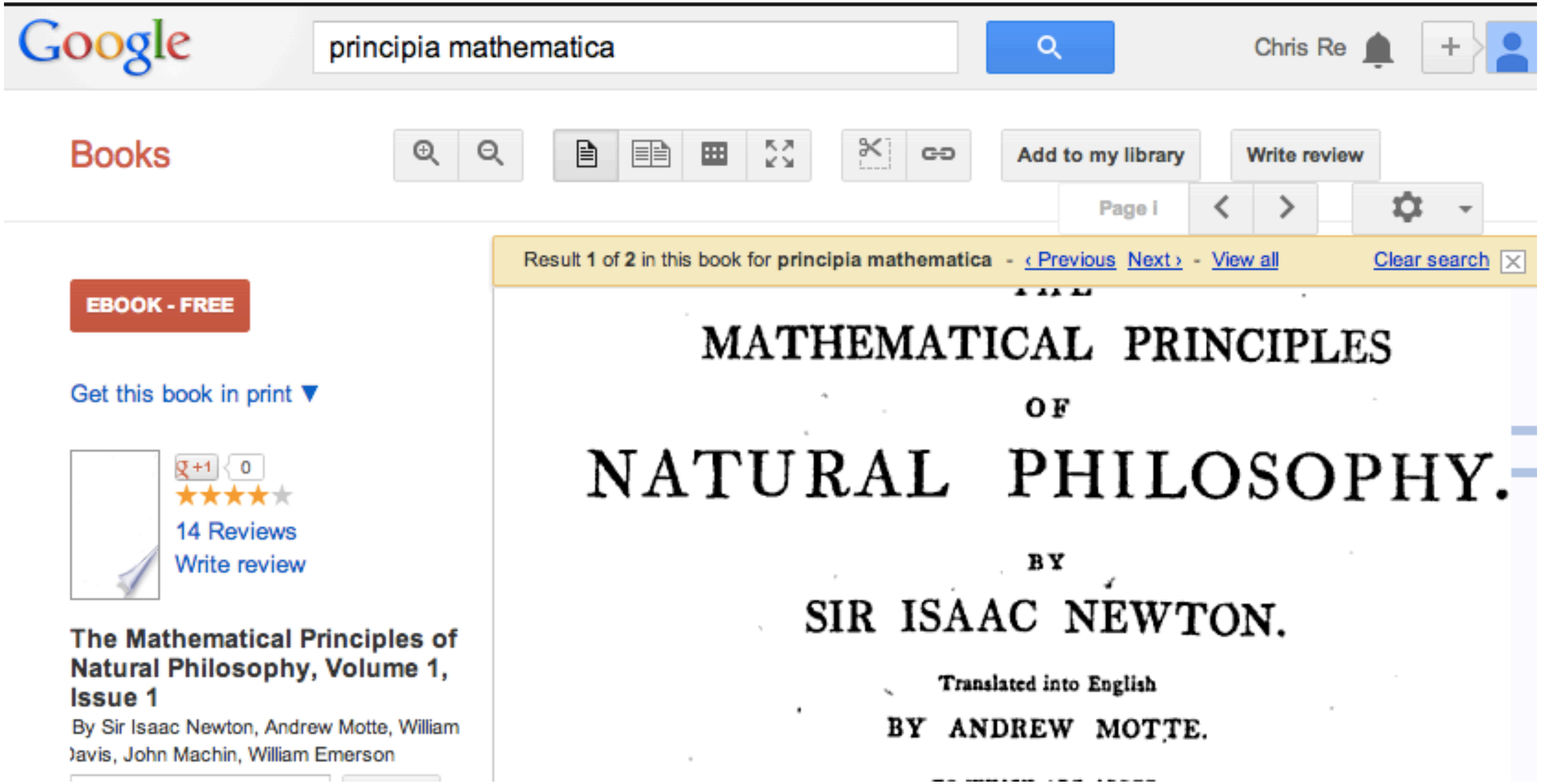


Principia
Mathematica

Turns out,
Newton was ***not*** alone

Millennia of Knowledge in Libraries

Now: freely available & digital



The world's scientific knowledge is **accessible**.

But we’re still human…

The world's scientific knowledge is **accessible**, but not **readable**.

Today, some pressing problems require **macroscopic knowledge**

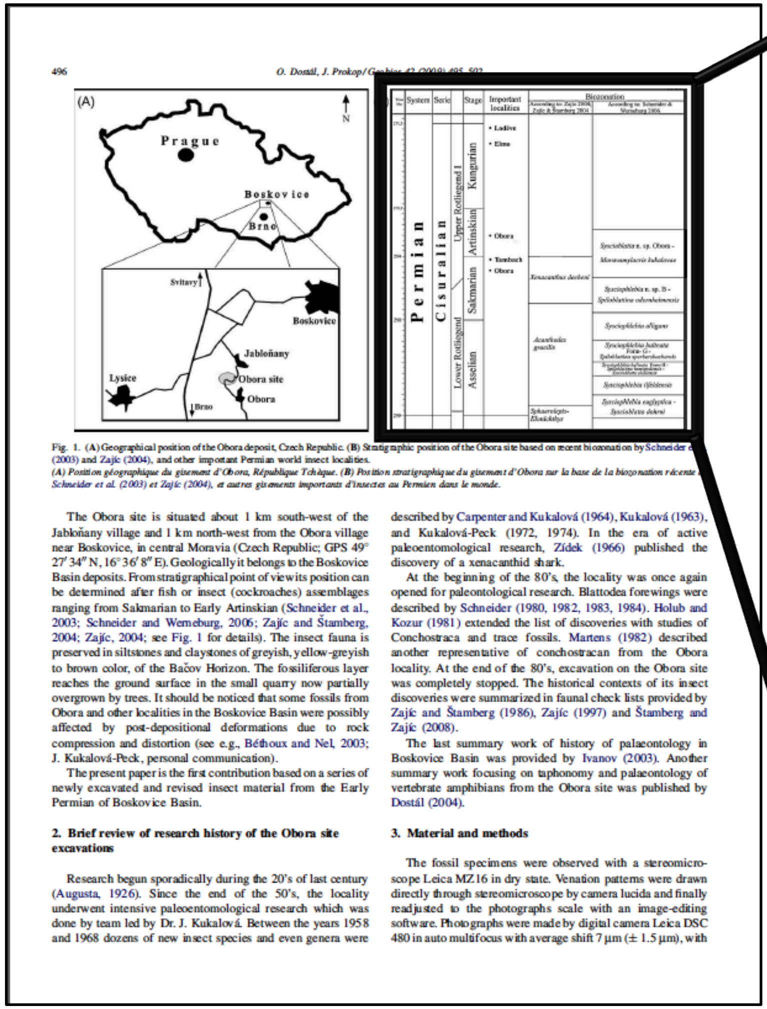
Climate & Biodiversity

Health

Financial Markets

Could we build a machine
to **read** for us?

Data are buried in tables, but not in a self-contained way



496 O. Dostál, J. Prokop / Geobios 42 (2009) 495–502

Fig. 1. (A) Geographical position of the Obora deposit, Czech Republic. (B) Stratigraphic position of the Obora site based on recent biozonation by Schneider et al. (2003) and Zajíc (2004), and other important Permian world insect localities.
(A) Position géographique du gisement d'Obora, République Tchèque. (B) Position stratigraphique du gisement d'Obora sur la base de la biozonation récente de Schneider et al. (2003) et Zajíc (2004), et autres gisements importants d'insectes au Permien dans le monde.

The Obora site is situated about 1 km south-west of the Jabloňany village and 1 km north-west from the Obora village near Boskovice, in central Moravia (Czech Republic; GPS 49° 27′ 34″ N, 16° 36′ 8″ E). Geologically it belongs to the Boskovice Basin deposits. From stratigraphical point of view its position can be determined after fish or insect (cockroaches) assemblages ranging from Sakmarian to Early Artinskian (Schneider et al., 2003; Schneider and Werneburg, 2006; Zajíc and Štamberg, 2004; Zajíc, 2004; see Fig. 1 for details). The insect fauna is preserved in siltstones and claystones of greyish, yellow-greyish to brown color, of the Bačov Horizon. The fossiliferous layer reaches the ground surface in the small quarry now partially overgrown by trees. It should be noticed that some fossils from Obora and other localities in the Boskovice Basin were possibly affected by post-depositional deformations due to rock compression and distortion (see e.g., Béthoux and Nel, 2003; J. Kukalová-Peck, personal communication).

The present paper is the first contribution based on a series of newly excavated and revised insect material from the Early Permian of Boskovice Basin.

2. Brief review of research history of the Obora site excavations

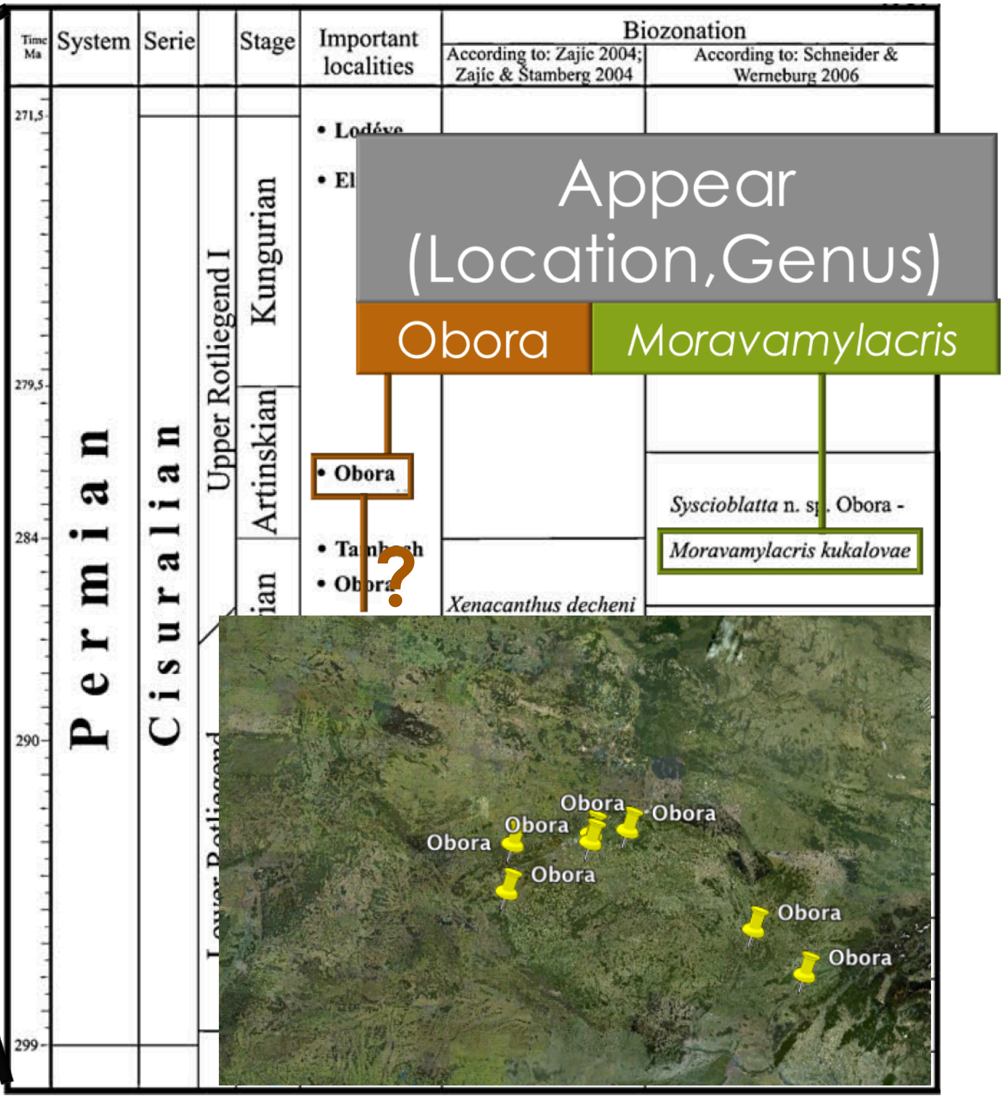
Research begun sporadically during the 20's of last century (Augusta, 1926). Since the end of the 50's, the locality underwent intensive paleoentomological research which was done by team led by Dr. J. Kukalová. Between the years 1958 and 1968 dozens of new insect species and even genera were described by Carpenter and Kukalová (1964), Kukalová (1963), and Kukalová-Peck (1972, 1974). In the era of active paleoentomological research, Zídek (1966) published the discovery of a xenacanthid shark.

At the beginning of the 80's, the locality was once again opened for paleontological research. Blattodea forewings were described by Schneider (1980, 1982, 1983, 1984). Holub and Kozur (1981) extended the list of discoveries with studies of Conchostraca and trace fossils. Martens (1982) described another representative of conchostracan from the Obora locality. At the end of the 80's, excavation on the Obora site was completely stopped. The historical contexts of its insect discoveries were summarized in faunal check lists provided by Zajíc and Štamberg (1986), Zajíc (1997) and Štamberg and Zajíc (2008).

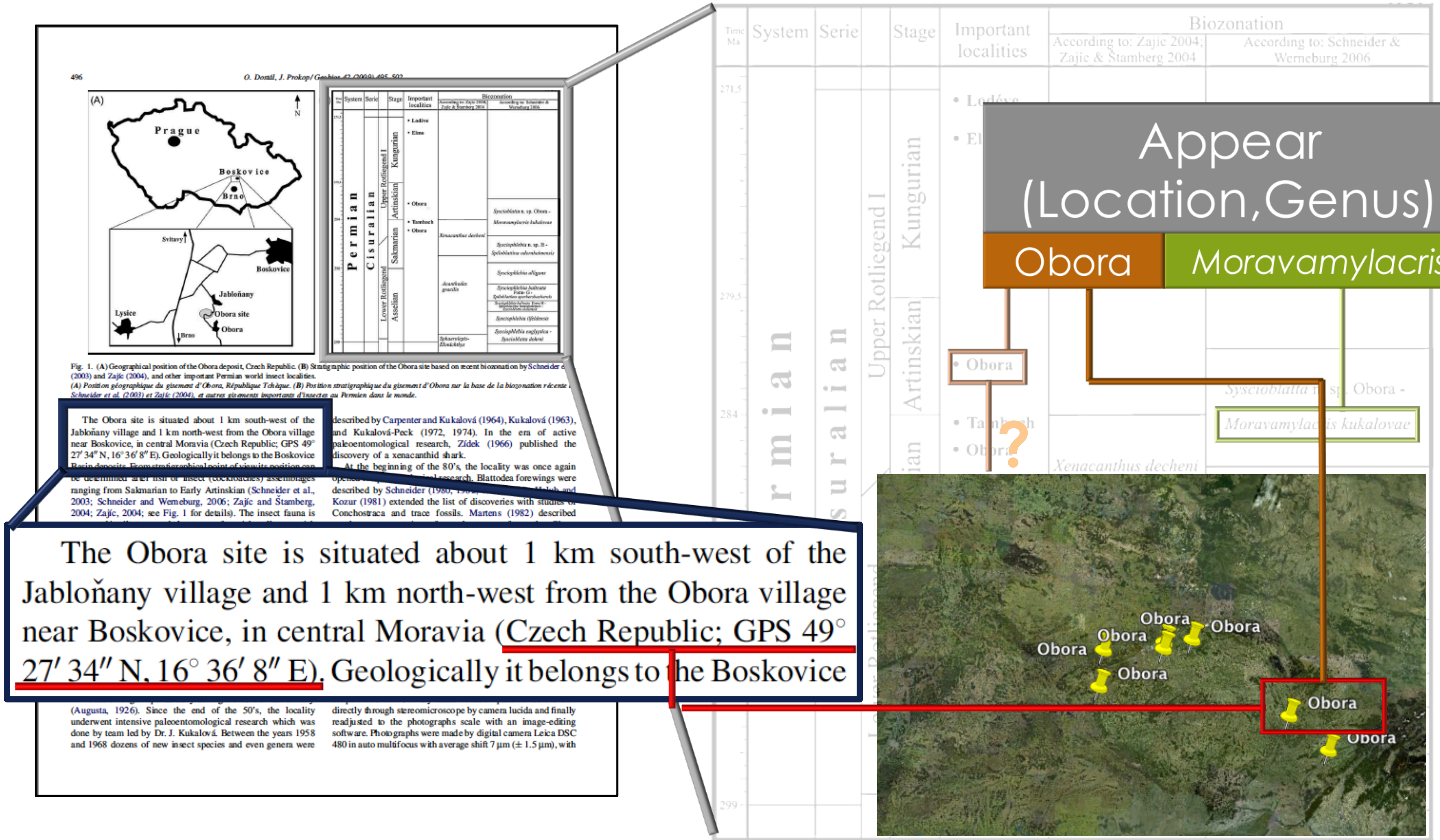
The last summary work of history of palaeontology in Boskovice Basin was provided by Ivanov (2003). Another summary work focusing on taphonomy and palaeontology of vertebrate amphibians from the Obora site was published by Dostál (2004).

3. Material and methods

The fossil specimens were observed with a stereomicroscope Leica MZ16 in dry state. Venation patterns were drawn directly through stereomicroscope by camera lucida and finally readjusted to the photographs scale with an image-editing software. Photographs were made by digital camera Leica DSC 480 in auto multifocus with average shift 7 μm (± 1.5 μm), with



Data are buried in tables, but not in a self-contained way



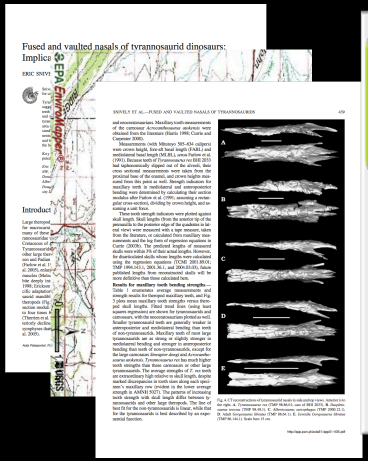
PaleoDeepDive

Shanan Peters (Geo) and Miron Livny (CS)
DeepDive.Stanford.edu (**Ce Zhang** et al.)

PaleoDeepDive

The Goal

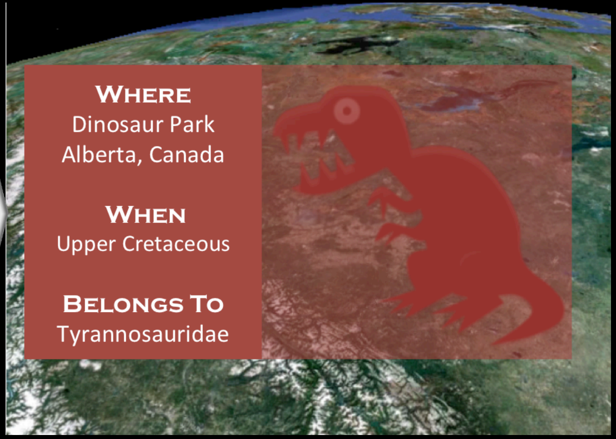
Extract paleobiological facts to build higher coverage fossil record.



T. Rex are found dating to the upper Cretaceous.

Statistical Inference

Appears("T. Rex", "Cretaceous")



Aggressive Approach

Every character, word, part of speech is a variable

statistical Inference *on billions of variables.*

nature International weekly journal of science

PaleoDB

Human-created

329 volunteers

13 years

46K documents

200+ Papers,
17 Nature/Science

PaleoDeepDive

Machine-created

10x **documents.**

100x **extractions.**

Preliminary Precision

Formation **Precision**	
	PaleoDB Volunteers: **0.84**
	PaleoDeepDive: **0.94**

Peters, S., Zhang, C, Livny, M., and Ré, C. A New Machine-Aggregated Empirical History of Life on Earth. *PLOS ONE, 2014 in Nature July 15*

Hope: knowledge bases can help **accelerate science**.

Tree of Life

Drug Repurposing

Genomics

Used by a number of companies.
and winner of TAC14 Slot Filling competitions.

Human Trafficking on the (Dark) Web…

Hypothesis: Trafficked individuals offer ***lower cost*** and ***riskier*** sexual services.

In Plain sight: Web ads for such services

Challenges:

1. Need **high-resolution information** to build model.
   - *services for what rate, ethnicity, location, etc.*
2. Scientific papers are **clear**—*dark web is* **obfuscated**.

Human Trafficking on the (Dark) Web…

DARPA MEMEX

Web Text

NASA NYU VIDA Where data comes to life

DeepDive

Structured Info: Phone #, Rates of Service, …

Normal two call agency. Jessica was ready when I arrived at body. A couple of tattoos but not in visual-annoying locations pleasing the client - definitely not GFE. She does the basics. recommending - I may repeat but only because of her looks.

Drug_Abuse?
Forced_Prostitution?

Use: *law enforcement/ NGOs*

In Use by Law Enforcement

New York DA use MEMEX Data for all trafficking investigations this year. ***Real Arrests***

For DARPA MEMEX, we were operational in 6 months

- Processed >35M documents (~26M records)
- Tens of columns (location, phone #, price, etc)
- With compute times of less than a day
- >90% Precision for most relations

What you will learn about in this section

1. A Crash Course in Analytics

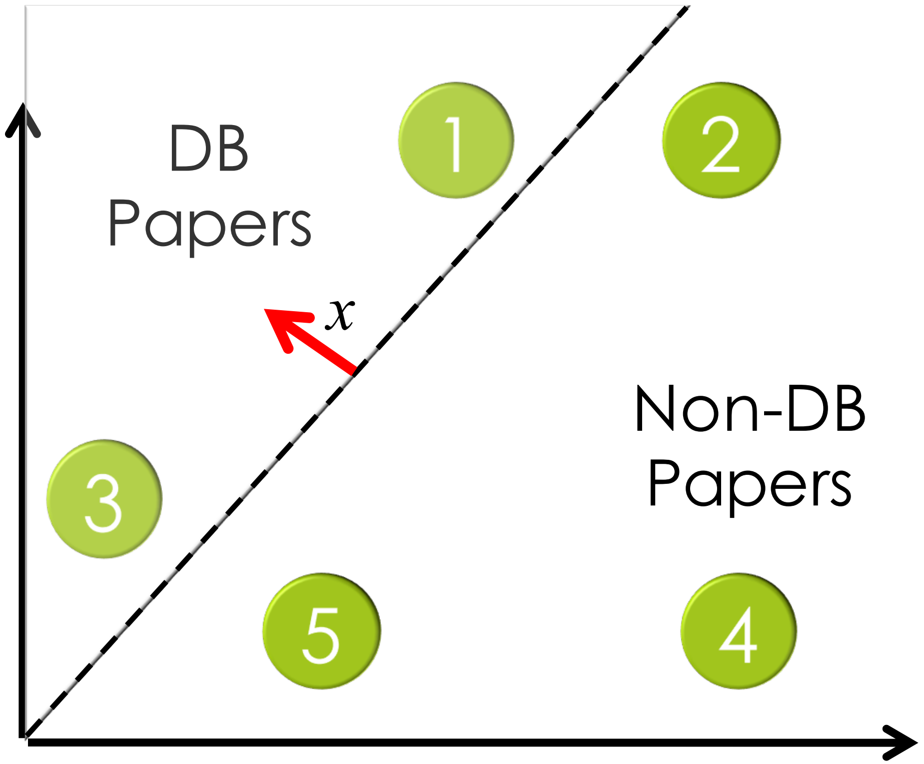
2. ACTIVITY: Run SGD!

Example: Linear Models

Label papers as DB Papers or Non-DB Papers

1. Map papers to R^d

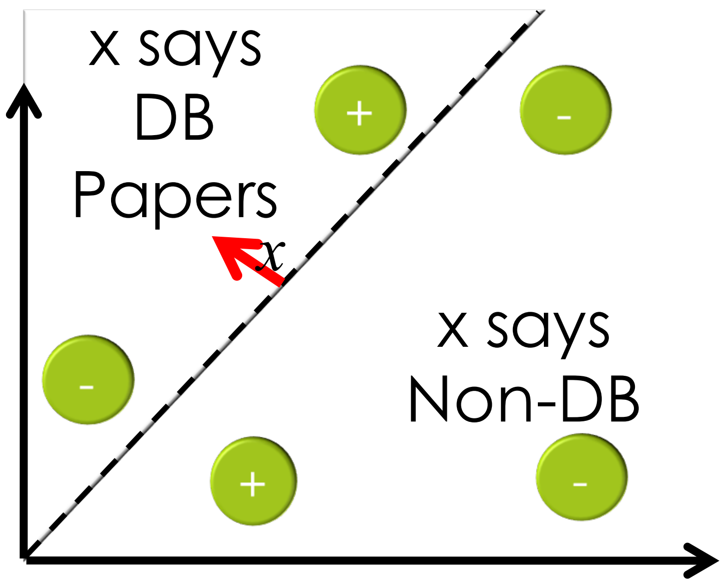
2. Classify via plane



How do we pick a good plane, x?

Example: Linear Models

Input: *Labeled* papers.
Each point labeled as
DB paper (+) or not (-)



Idea: score each plane

$$\min_x \sum_{i=1}^{N} f(x, y_i)$$

y_i is a paper vector and its label

f is convex
function in x

*e.g. squared distance (**least squares**), hinge loss (**svm**),*
*log loss (**logistic regression**)*

Separable Inverse Problems

$$\min_x P(x) + \sum_{i=1}^{N} f(x, y_i)$$

x	the model
y_i	a data item
f	scores the error
P	Enforces prior

Paper Classification: y_i is the paper with its label, and x is a vector of weights

Neutrinos: x is neutrino path and y_i is a DOM reading

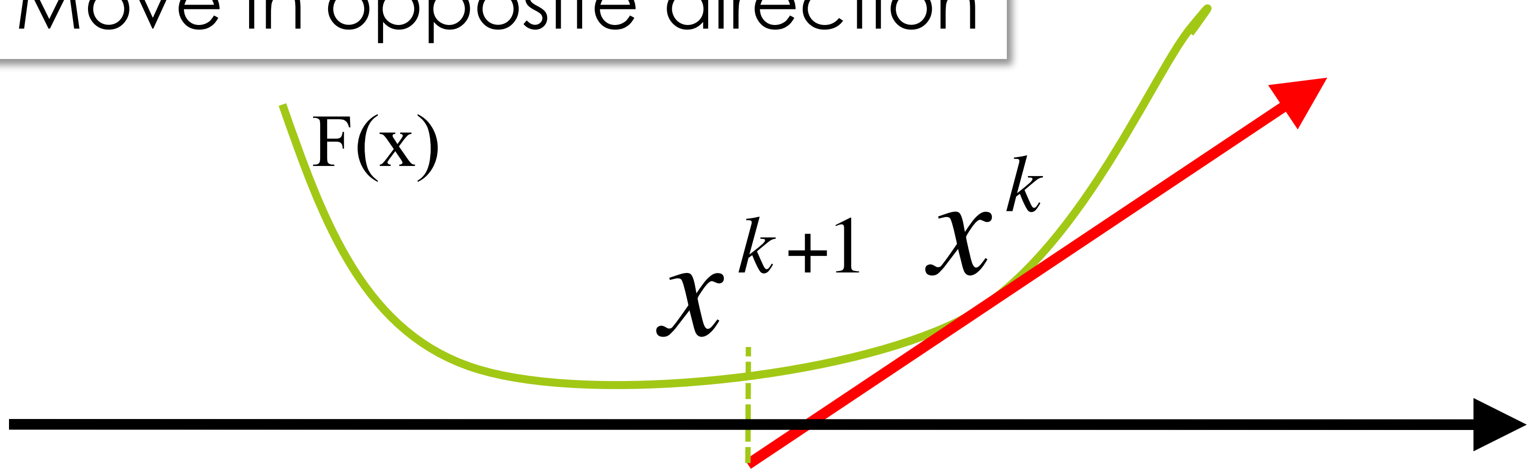
Conditional Random Fields, **Low-rank factorizations (relaxations)**, **Logistic regression**, **Lasso**, **etc.**

Claim: General data analysis technique that is no more difficult to compute than SQL AVG

Background: Gradient Methods

$$F(x) = \sum_{i=1}^{N} f(x, y_i) \qquad x^{k+1} = x^k - \alpha \nabla F(x^k)$$

Gradient Methods: Iterative
1. Start at current x^k,
2. Take gradient at x^k, and
3. Move in opposite direction



Background: Incremental Gradient Methods

[Robbins & Monro, 1951]

$$F(x) = \sum_{i=1}^{N} f(x, y_i)$$

Gradient Methods:

1. Start at current x,
2. ***Approximate*** gradient at x by selecting *j in [N]*

$$G(x, y_j) = N\nabla f(x, y_j) \qquad \nabla F(x) \approx G(x, y_j)$$

3. Move in opposite direction

$$x^{k+1} = x^k - \alpha G(x^k, y_j)$$

Single *data item* to approximate gradient

Incremental Gradient Methods

Why use iGMs? iGMs converge to an optimal for many problems, but the real reason is:

iGMs are *fast.*

$$x+ = -\alpha G(x^k, y_j)$$

Technical Connection:
iGM isomorphic to computing a SQL AVG

Solve statistical models in an RDBMS for free

IO Aware iGMs

iGMs involve cheap numerical computations, but require reading / writing a lot of data…

$$x+=-\alpha G(x^k, y_j)$$

We need to use our IO aware cost models to think about performance here!

Activity 19-1

iGMs ≈ User Defined Aggregate (UDA)

RDBMS UDA defined by 3 Functions

1. Initialize (State)
2. Transition(State, data)
3. Terminate(State)

State in $\mathbf{R}^2$: (# terms, total)
Transition($(n, T), d$) = $(n, T) + (1,d)$
Terminate((n,T)) = T/n

AVG

State in $\mathbf{R}^d$: *model* x
Transition(x, y_j) = $x - \alpha G(x,y_j)$
Terminate(x) = x

iGM

Widely used…

iGM can be integrated into existing databases for free.

Significance: Reuse DB internals for shuffling, compression, shared-nothing parallelism, and shared-memory allocators.

What you will learn about in this section

1. Trends that impact Analytics

2. Multicore

3. NUMA

4. SIMD

To get rid of **algorithms**,
we need a **fast** engine....

Modern **hardware** is amazing.

Message

Statistical algorithms *have* ***relaxed*** *notions of correctness; leads to* ***new opportunities*** *for:*

- Algorithms,
- Systems, and
- Hardware.

The Key Balance

DIMMWitted

Key Issue: Balance

<u>Statistical</u> versus <u>Hardware</u>

Efficiency.

Ce Zhang *DimmWitted: A Study of Main-Memory Statistical Analytics.* VLDB14.

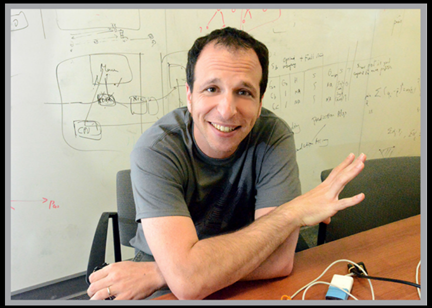
Modern **hardware** is amazing in 3 major trends for analytics:

(1) **Multicore**,
(2) **NUMA**, and
(3) **SIMD**.

Message: **Relaxing consistency** allows major performance improvements

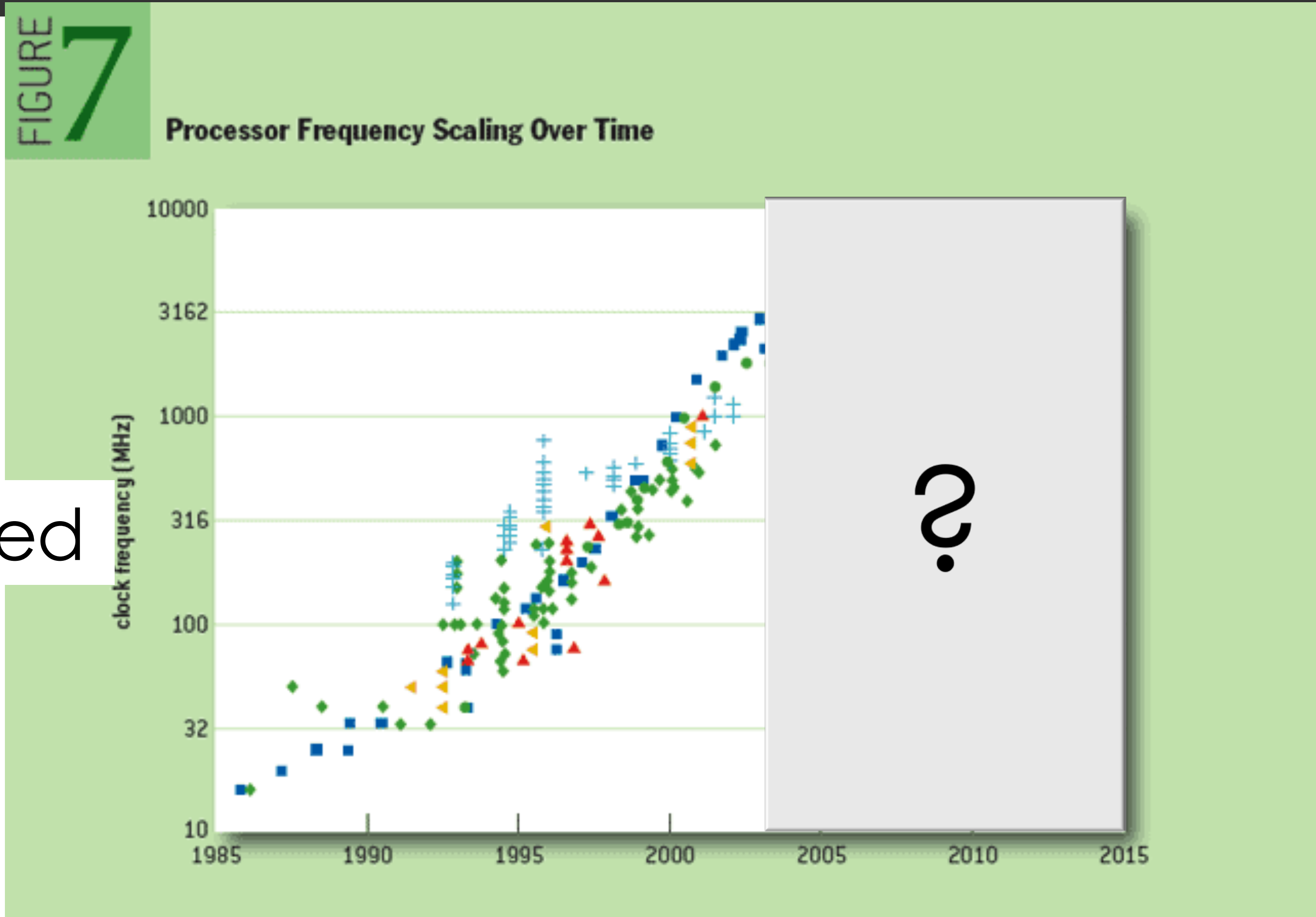
Trend 1: **Multicore**

[NIPS11]



Single Cores are not getting faster.

Speed



CPUDB: Danowitz et al. CACM QUEUE 2012

Chips contain multiple cores, so throughput is increasing… *but need to rewrite algos!*

Statistical Analytics Crash Course

Staggering amount of machine learning/stats can be written as:

$$\min_x \sum_{i=1}^{N} f(x, y_i)$$

N (number of y_is, data) typically in the billions

Ex: Classification, Recommendation, Deep Learning.

De facto iteration to solve large-scale problems: **SGD**.

$$x^{k+1} = x^k - \alpha \boxed{N \nabla f(x^k, y_j)}$$

Select one term, j, and estimate gradient.

Billions of tiny iterations.

How do we run SGD in Parallel?

Data Systems Perspective of SGD.

$$x^{k+1} = x^k - \alpha N \nabla f(x^k, y_j)$$

Insane conflicts: Billions of tiny jobs (~100 instructions), RW conflicts on x, which is called **the model.**

How can we hope to speed this up with parallelism? Serializability?

Multiple cores need locks to communicate!

How do we run SGD in Parallel?

High-level idea: Go Hogwild! (comment out the locks!) answer is only *statistically* correct.

Thm: If we do ***no locking***, SGD still converges to right answer—at essentially the same theoretical rate!

Hogwild! [Niu, Recht, **Ré**, Wright NIPS11]
AsySCD [Liu, Wright et al. ICML14, JMLR14]
Buckwild! [De Sa, Olukotun, Ré NIPS15]

Technical conditions on ratio of processors, delays, (semantic) sparsity.

Cortana: Microsoft's Digital Assistant

AI breakthrough: Microsoft's 'Project Adam' identifies dog breeds, points to future of machine learning

All web companies have similar: image rec, voice, mobile, search, etc.

"…using a technology called, of all things, ***Hogwild!****"*

http://www.wired.com/2014/07/microsoft-adam/
http://www.geekwire.com/2014/artificial-intelligence-breakthrough-microsofts-project-adam-identifies-dog-breeds/

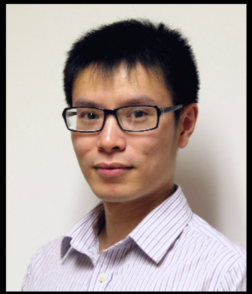
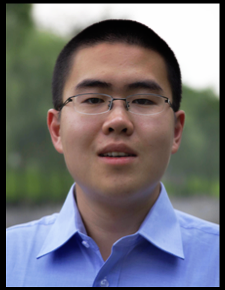
A larger trend?

Relaxing **consistency** to be **architecturally aware** can be a big performance win.

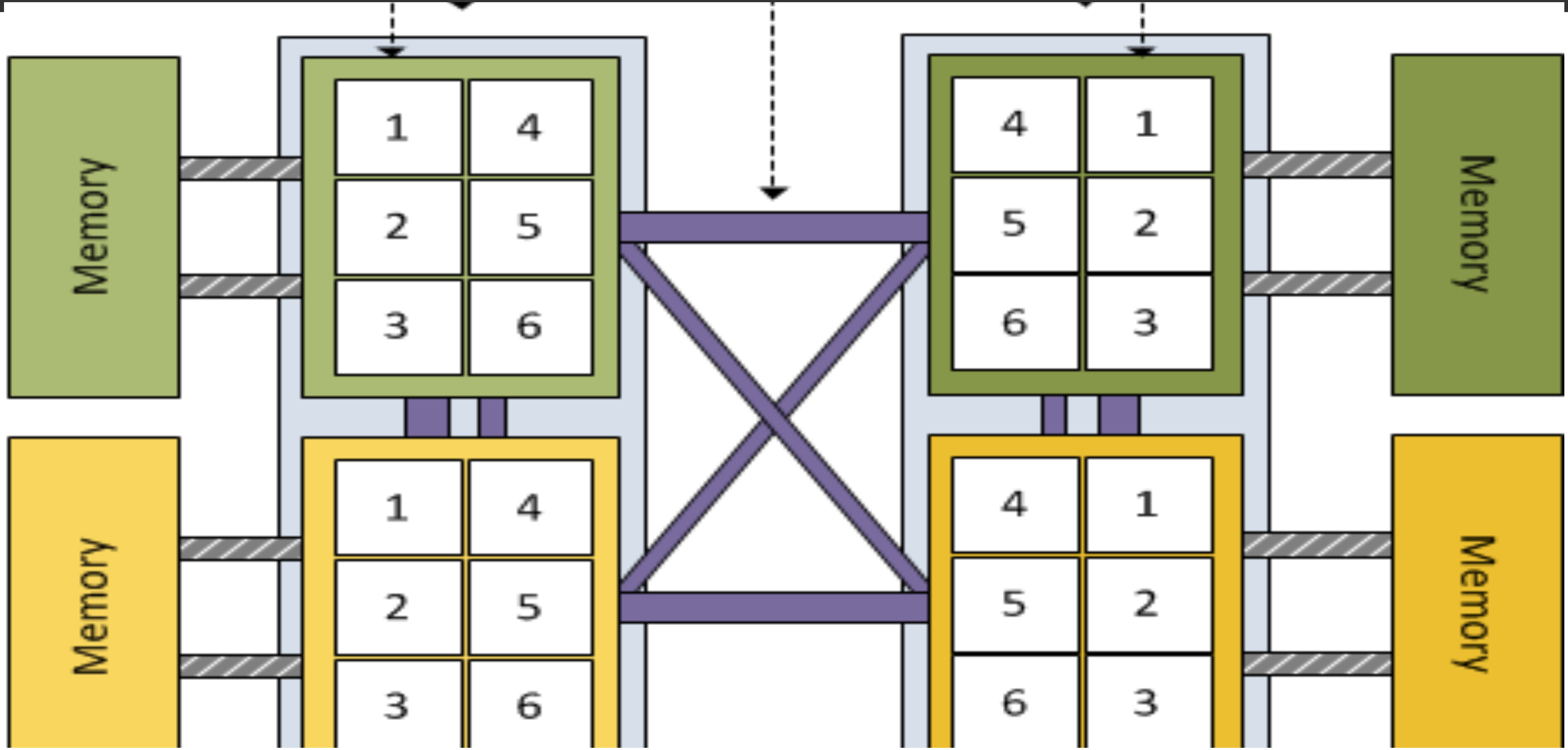
A larger trend?

Relaxing **consistency** to optimize around **architecture-specific IO cost models** can be a big performance win.

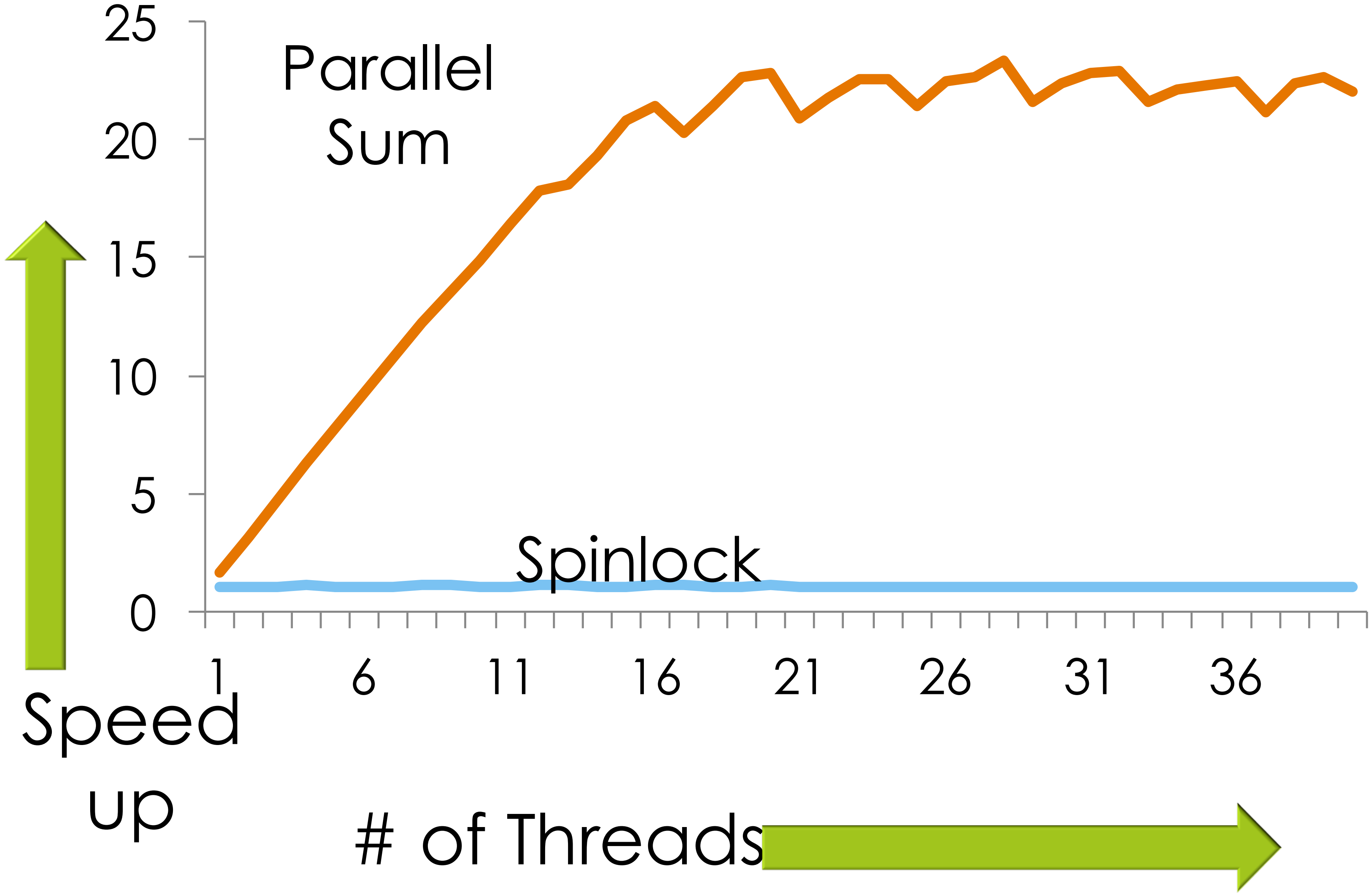
Trend 2: **NUMA**
[VLDB15, ICML14]

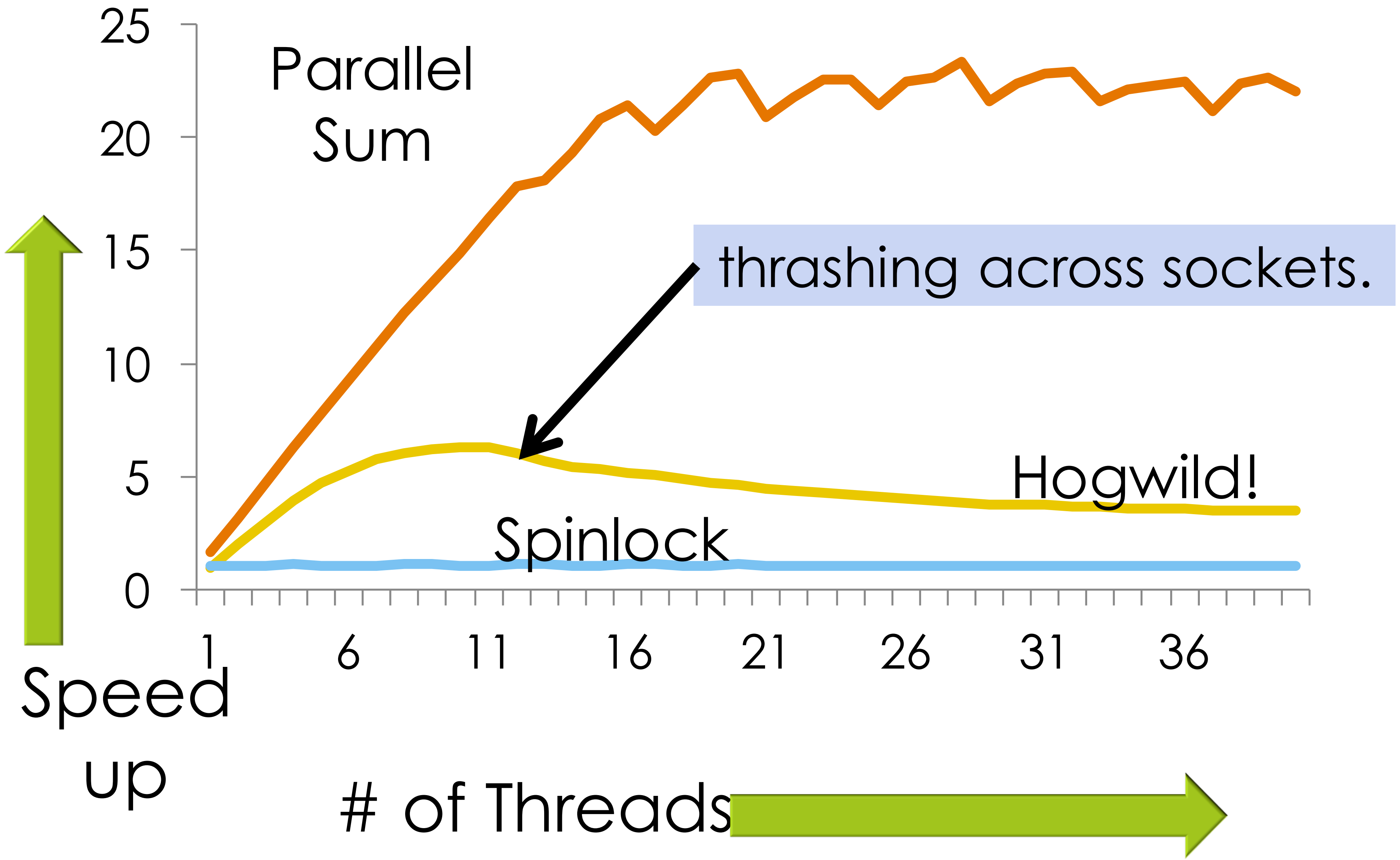
Simplified 4 socket.



One Example: Quadratic Programming with Orthant Constraints.

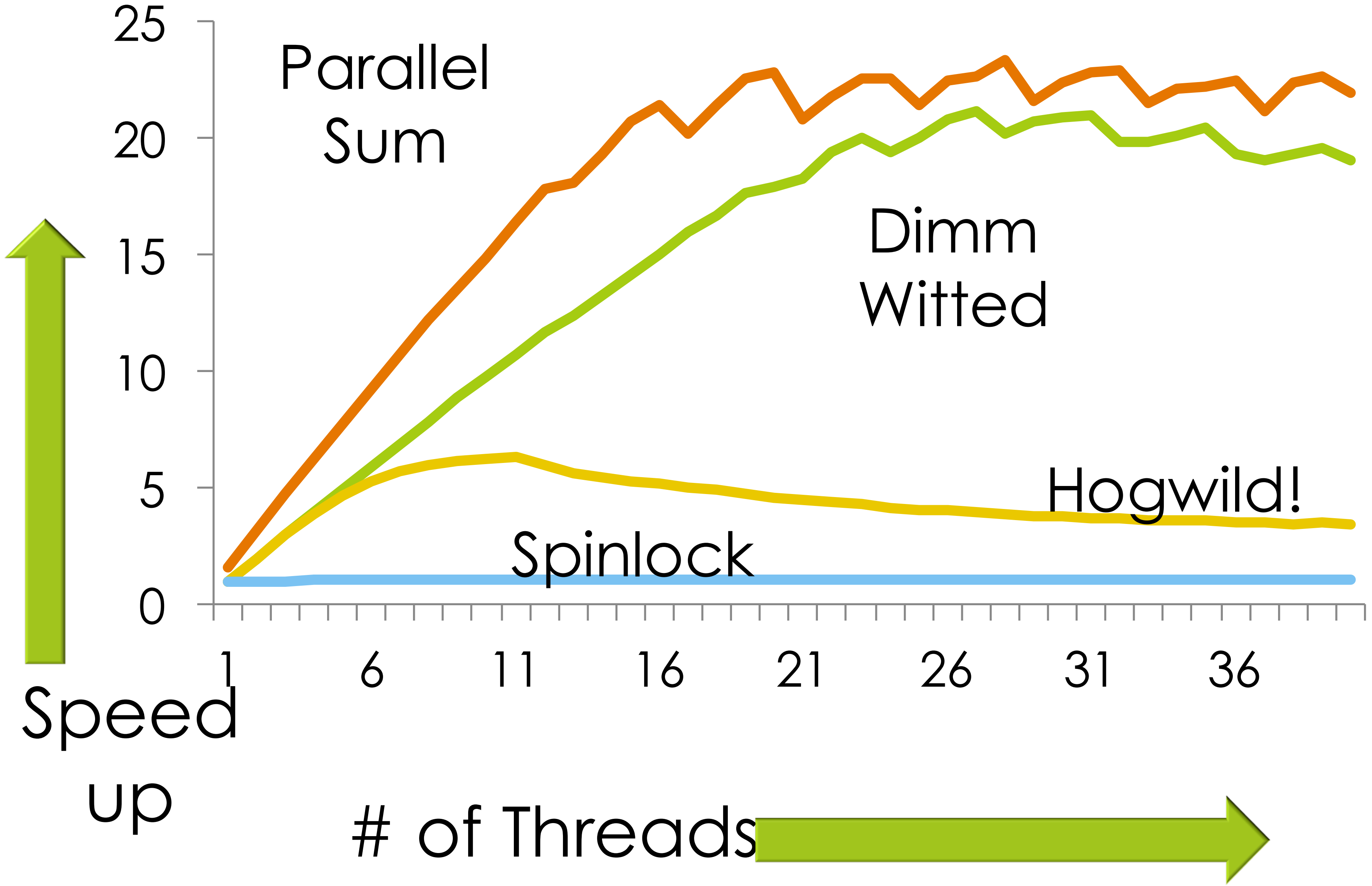


One Example.



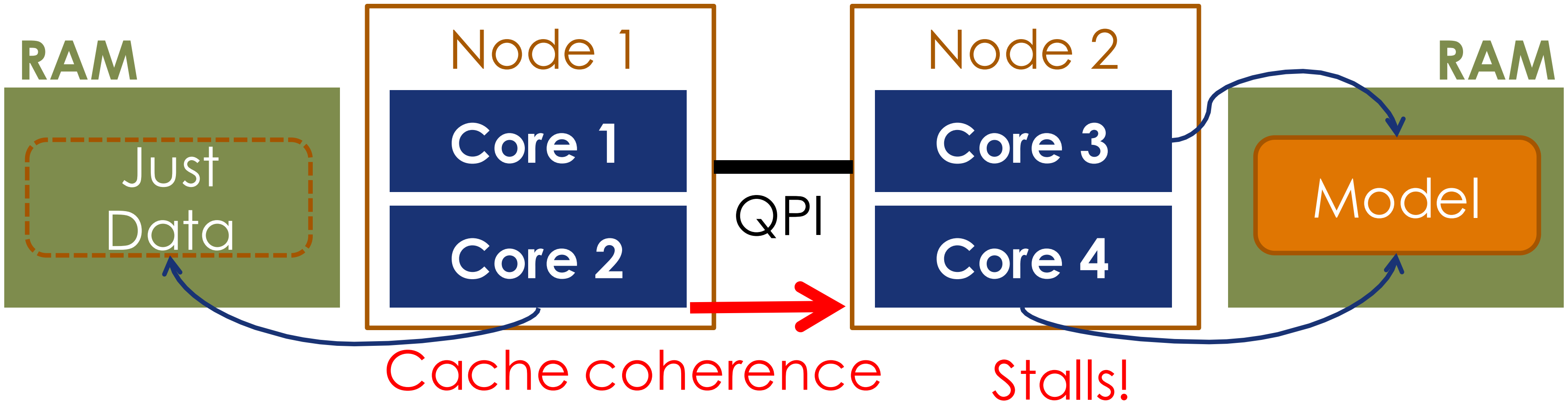
25
20
15
10
5
0
Parallel Sum
thrashing across sockets.
Hogwild!
Spinlock
1
6
11
16
21
26
31
36
Speed up
# of Threads

What about multiple sockets?

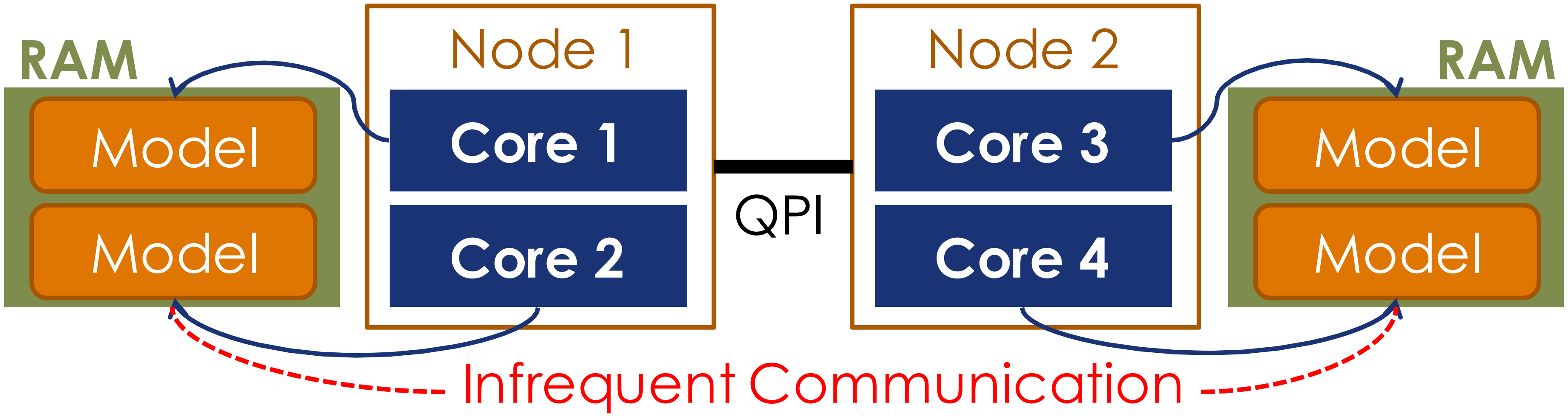


Model Replication

PerMachine (Hogwild!)



PerCore

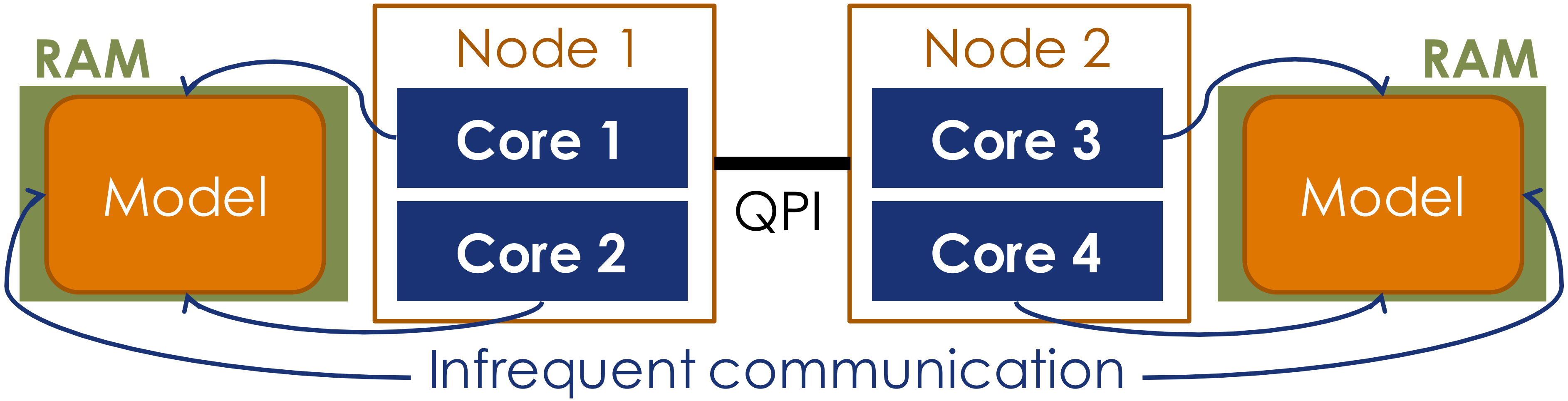


Model Replication

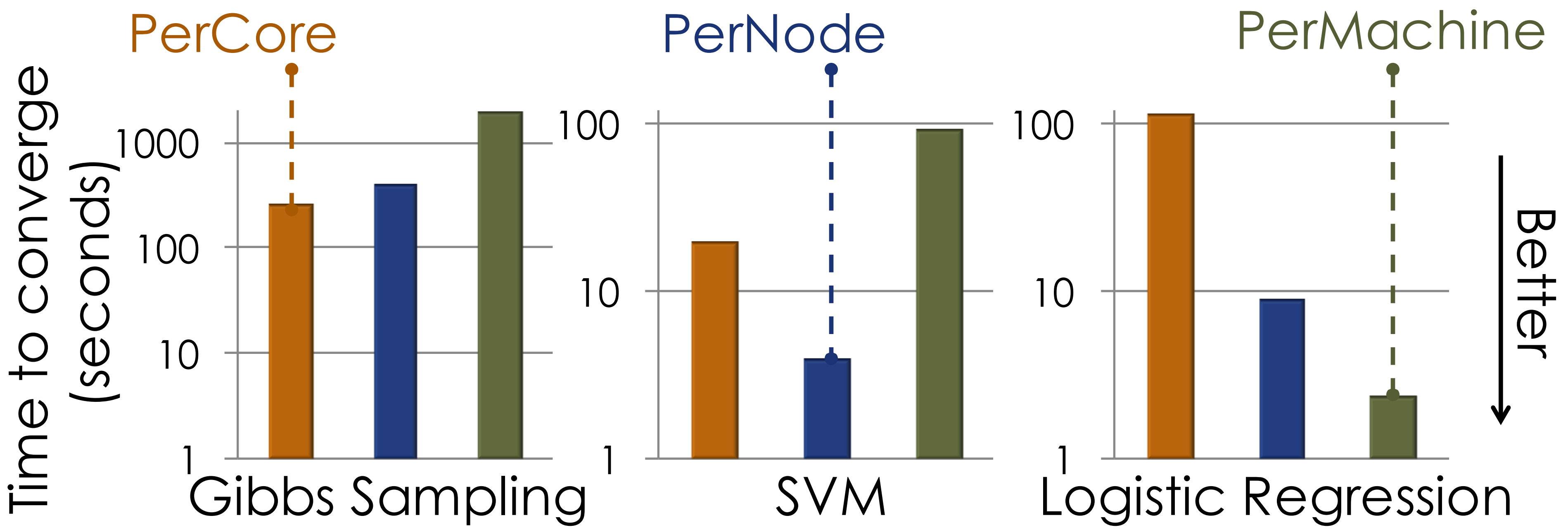
	Statistical Efficiency	Hardware Efficiency
Hogwild!	High	Low
PerCore	Low	High

In between both **Hogwild!** and **PerCore**?

PerNode



Model Replication Results



(1) No one method dominates the other.
(2) Each can be significantly better.

Statistical versus Hardware Efficiency

Relaxing consistency results in new tradeoffs.

1. Access methods
   - {Row, Column, Row-col}
2. Model Replication
   - {Core, Node, Machine}
3. Data Replication [SODA16]
   - {Full, Importance, Shard}

Can be 100x faster than classical choices, e.g., MLlib or GraphLab

DimmWitted: A Study of Main-Memory Statistical Analytics. VLDB14.

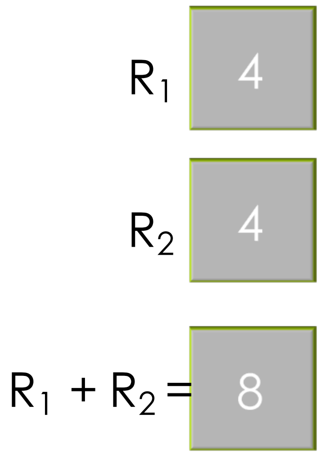
Trend 3: SIMD [NIPS15]

Modern processors offer **fine-grained** parallelism.

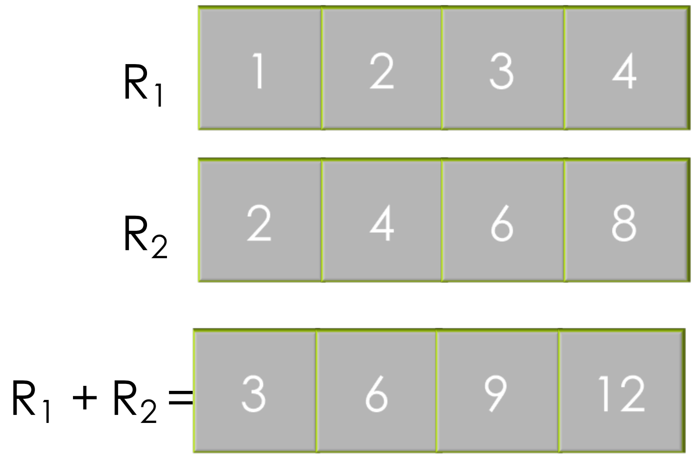
SIMD Processing: Fine-grained parallelism

Single instruction multiple data (SIMD)

Standard Addition



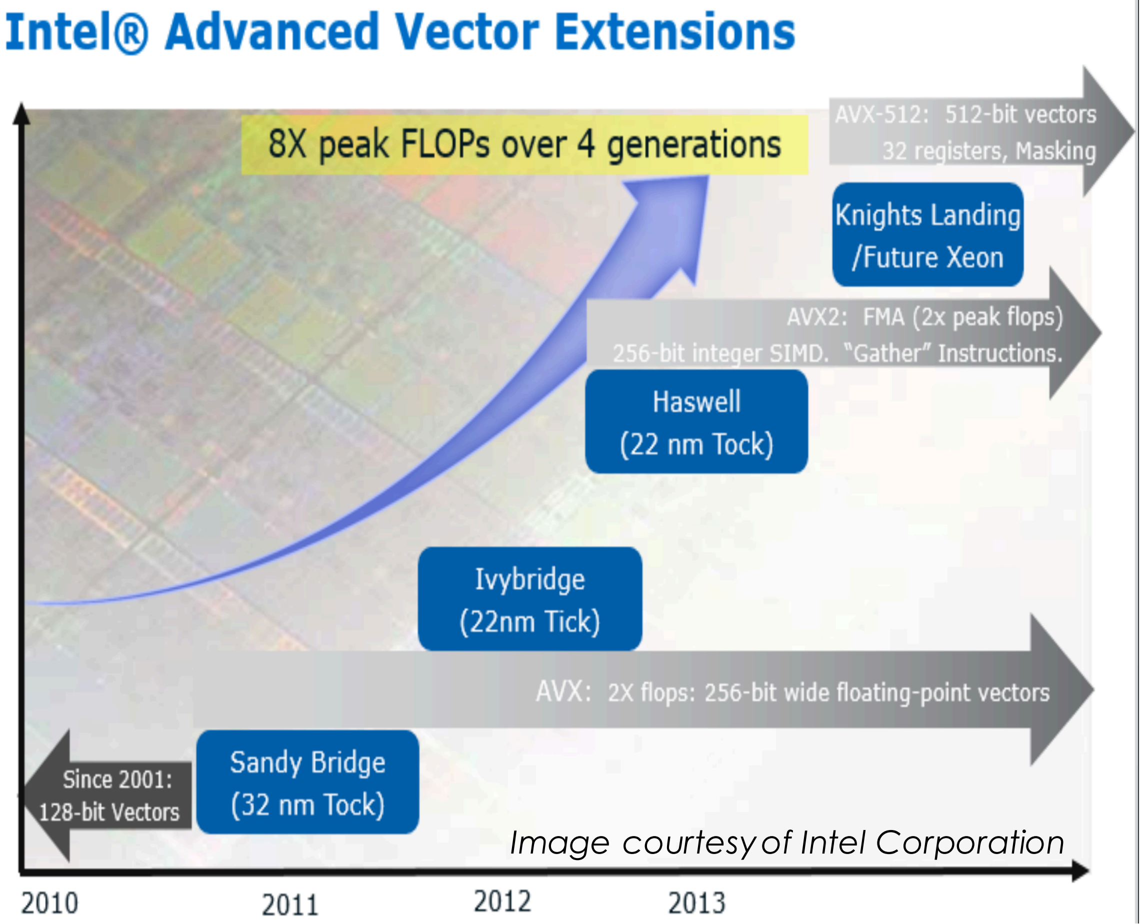
SIMD Addition (4 way)



Same operation on
multiple data points in parallel

SIMD Trend: Doubling again!

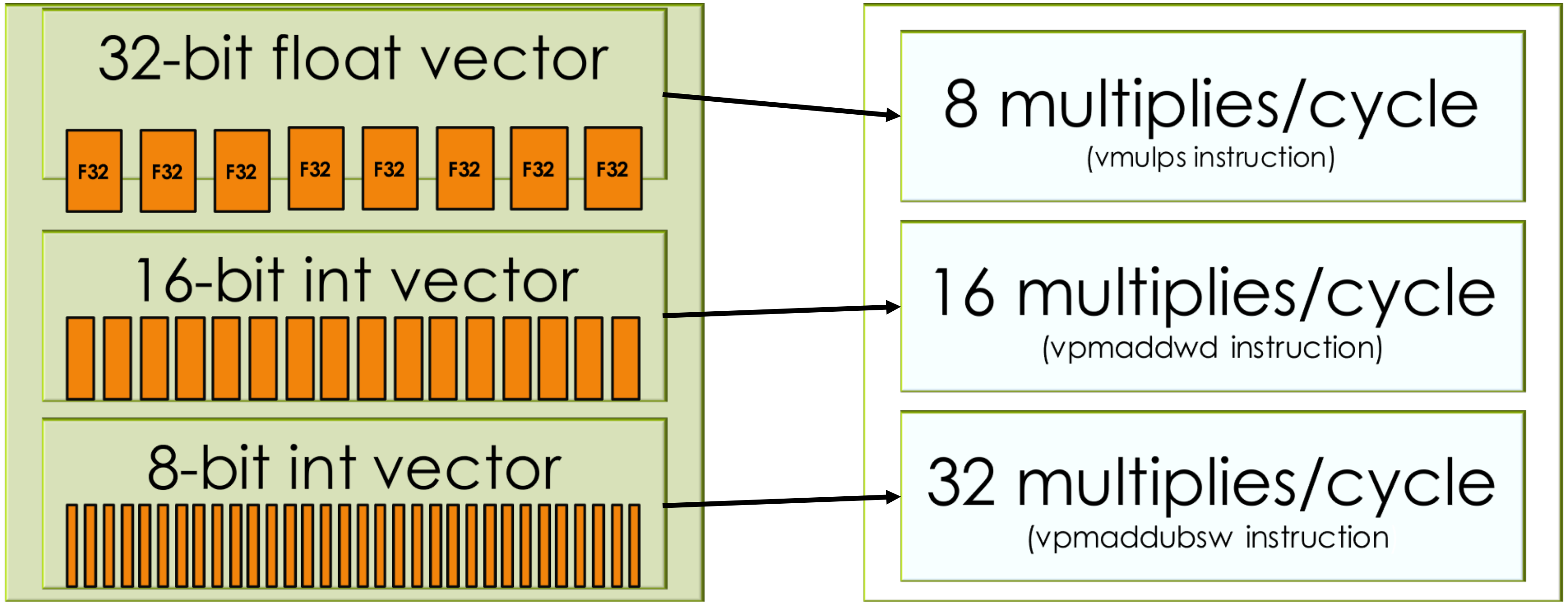
SIMD bandwidth has **doubled each** of the last four generations.



Good old days of Moore's Law! …
*If we can take advantage of **parallelism***

Precision vs. Parallelism

SIMD Precision | **SIMD Parallelism**



Tradeoff between precision & parallelism

BUCKWILD! [NIPS15]

BUCKWILD! Is an asynchronous SGD that **down-sample precision of data**

Statistical theory:

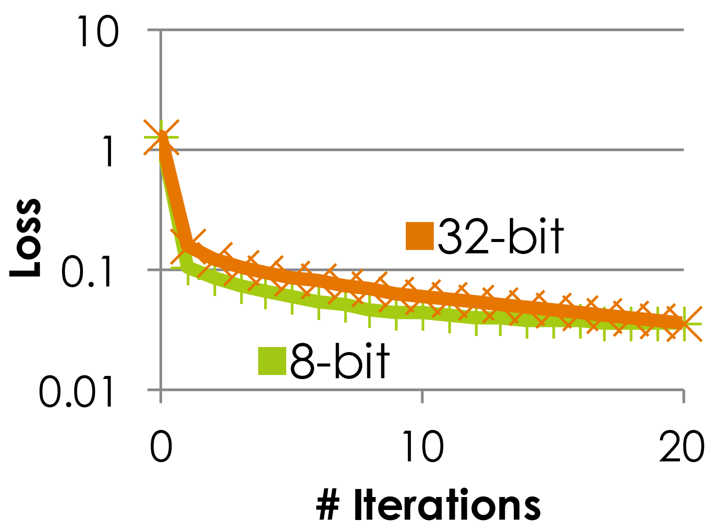
- Low-precision arithmetic essentially **no additional error**
  - when round-off error below the **noise floor** from the stochastic sampling.

Systems:

- Low-precision **increases effective memory bandwidth**
- It also decreases the load on the cache hierarchy.
- Near-linear compute speedup from SIMD.

Statistical vs. Hardware Efficiency

Same statistical efficiency!



Improved hardware efficiency!

- 8-bit gives about 2.5x speed up!
- Also, we can scale up to more threads without becoming memory bound

BUCKWILD! same **statistical efficiency**, while greater **hardware efficiency**.

Conclusion

1. **Statistical Algorithms** *offers new opportunity*

2. **Hardware v. Statistical** Efficiency.

Conclusion

We can trade off ***statistical efficiency*** to focus on optimizing our IO cost (***hardware efficiency***)!