Trends in Scientific Computing

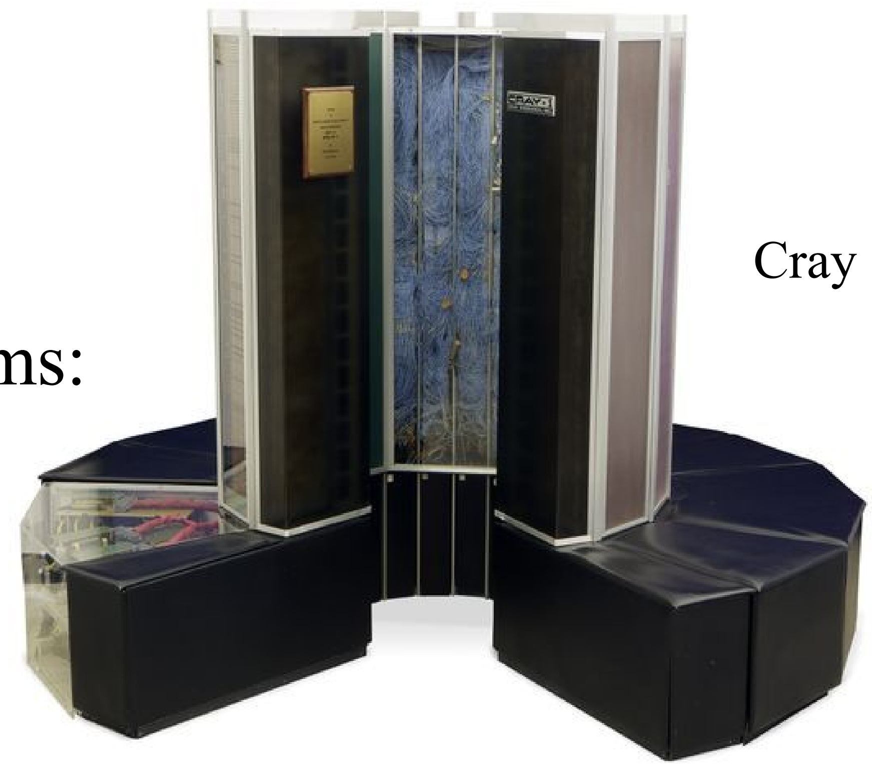
Timothy A. Gonsalves
IIT Mandi

26th Nov 2018

Scientific Computing: in the Beginning

Specialised hardware from Cray *et al*.

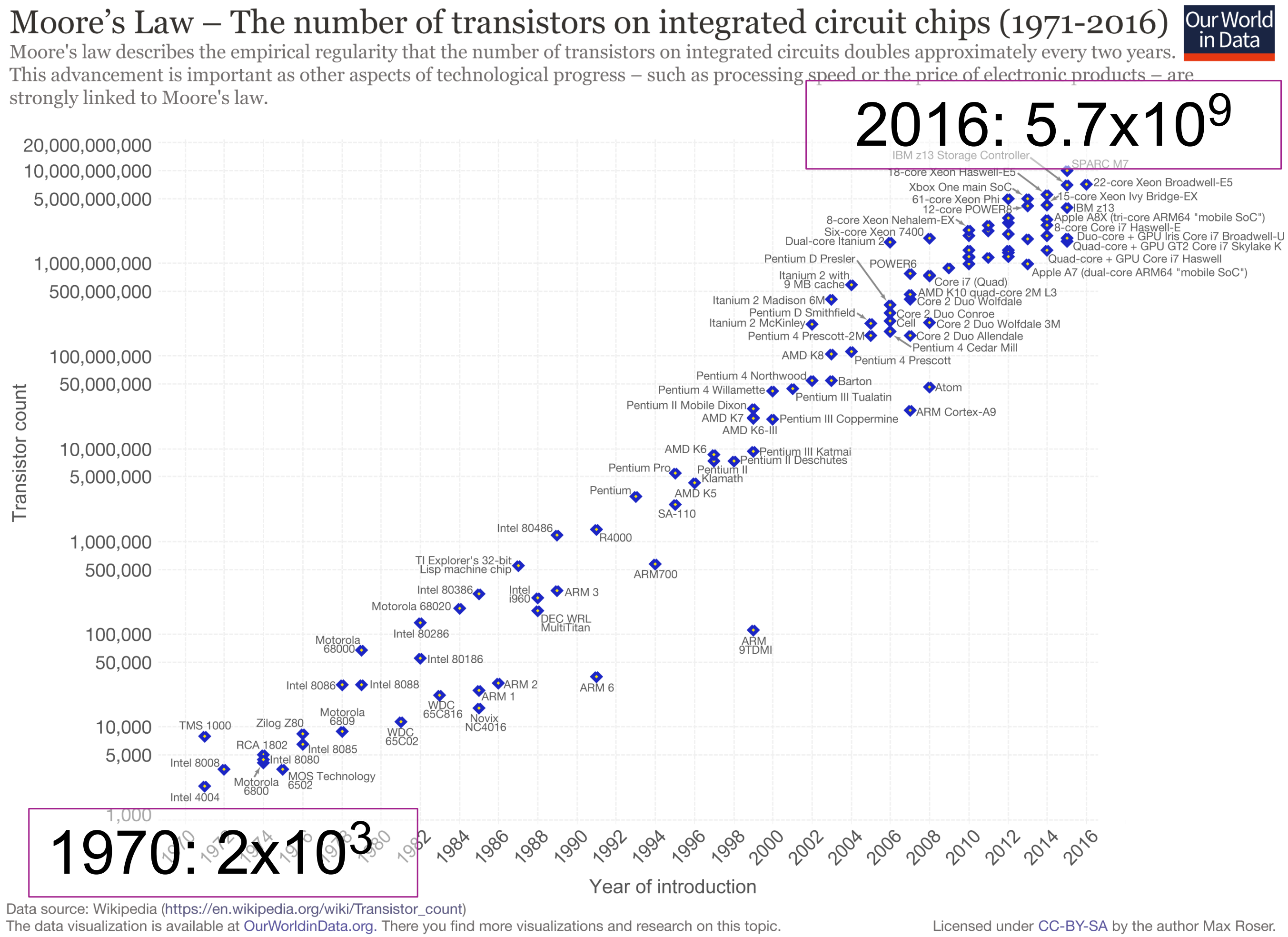
- Vector processors – multiple arithmetic units
- Multiple ALUs
- Large shared memory
- Very expensive
- Parallel Fortran
- Suitable for regular problems: matrix manipulations, etc



Cray 1

Fundamental Laws (informally)

1. Moore's Law: the performance of computer hardware doubles every 1.5 years



[Wikipedia]

Communications

Today

- Near universal availability of intermittent shared wireless at 10s-100s kb/s
- Network failures and congestion are common

2035

- Universal availability of shared wireless at 1-100 Mb/s
- Congestion and failures will still cause “intermittent” behaviour for high-bandwidth uses

Personal Hardware

Parameter	2015	2035*
Pocket-sized device	Lava 356 phone	XYZ digital assistant
CPU	1 core, 1 GHz	8,000 core, 1 GHz
RAM	0.5 GB	4 TB
Storage	32GB	128 TB
Network (wireless from telco)	100 kb/s shared	100 Mb/s shared
Display	2”x2”	10”x10” foldable
Security	PIN (insecure)	Biometric (unbreakable)
Cost (in 2015 Rs.)	Rs. 4,500	Rs. 2,000

2035: Cradle-to-grave pocket device = 8,000 today's laptops

* Based on Moore's Law

Evolution of Computer Hardware

- Moore's Law + Economies of Scale + PC Revolution
- ==> Design HPC using commercial off-the-shelf (COTS) microprocessors
- CPU power of N microprocessors >
  CPU of custom-built Cray
  AND
  Cost of COTS << Cost of Cray
- Shared memory multi-processors using COTS CPUs

… Evolution of Computer Hardware

- Shared memory multi-processors using COTS CPUs
  - interconnection network is bottleneck
  - cost increases, performance decreases
- Need a change in programming style, moving away from highly parallelisable matrix computation

How to address the computational needs of scientists and engineers?

Fundamental Laws (informally)

2. Metcalfe's Law: the usefulness of a network is proportional to the square of the number of users

$$V_{Tencent} = 7.39 \times 10^{-9} \times n^2$$

$$V_{Facebook} = 5.70 \times 10^{-9} \times n^2$$

[Wikipedia]

3. Law of Large Numbers: behaviour of a large population tends to a predictable normal distribution

==> Machine learning AI

Technology Disruptions

IoT: Inexpensive sensors connected to Internet

Weather stations:

- IMD: Rs. 10s lakhs each, in major towns, district HQ
  - Highly accurate
- IoT: Rs. 1,000s each, in every farm or field
  - Less accurate
  - Correct using statistical software

Likewise, GPS, web cameras, motion sensors, ...

- Costs declined from Rs. lakhs to Rs. 1,000s
- Long-life with batteries/solar-power

$\Longrightarrow$ Vast quantities of data about every locality

… Technology Disruptions

Cloud Storage and Computing

- Easy, world-wide sharing of data

Data Analytics

- Vast amounts of data ==> use probability and statistics to make inferences

Machine Learning

- Given large number of case histories, ML algorithms identify patterns to make predictions
- Learns from experience

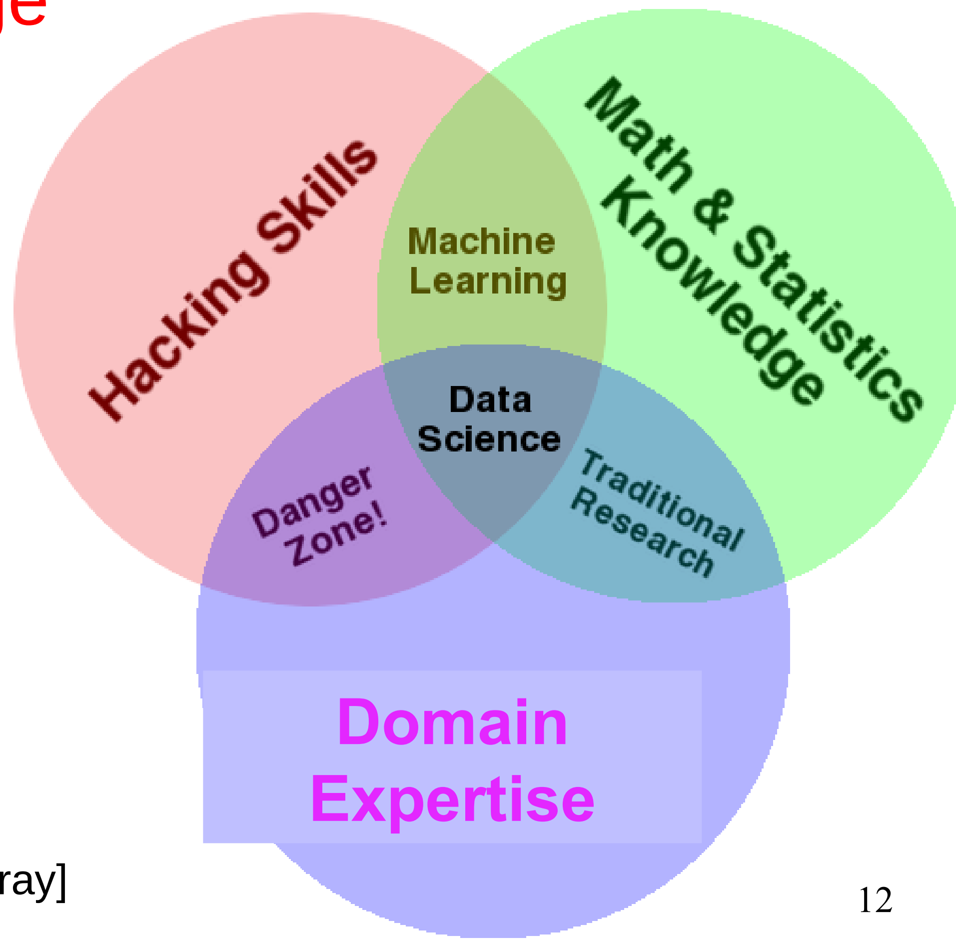
Evolution of Scientific Computing

- Two converging trends
  - Hardware tending towards very large clusters with non-uniform memory + cloud computing
  - Interesting scientific and engineering computation moving from solution of equations (regular) to searching for patterns in very large data sets (irregular)

    ==> Data Science

What is Data Science?

- Application of scientific processes, algorithms and systems to extract knowledge and insights from data in various forms [Wikipedia]
- Aspects of Data Science:
  - Data collection
  - Data analysis
  - Inference
  - Communication of results
- 4th paradigm of science: empirical, theoretical, computational, data science [Jim Gray]



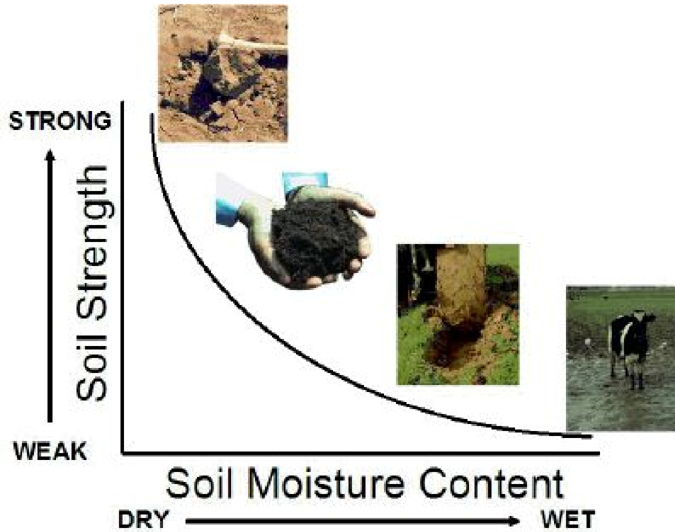
[Drew Conway]

Examples of Data Science

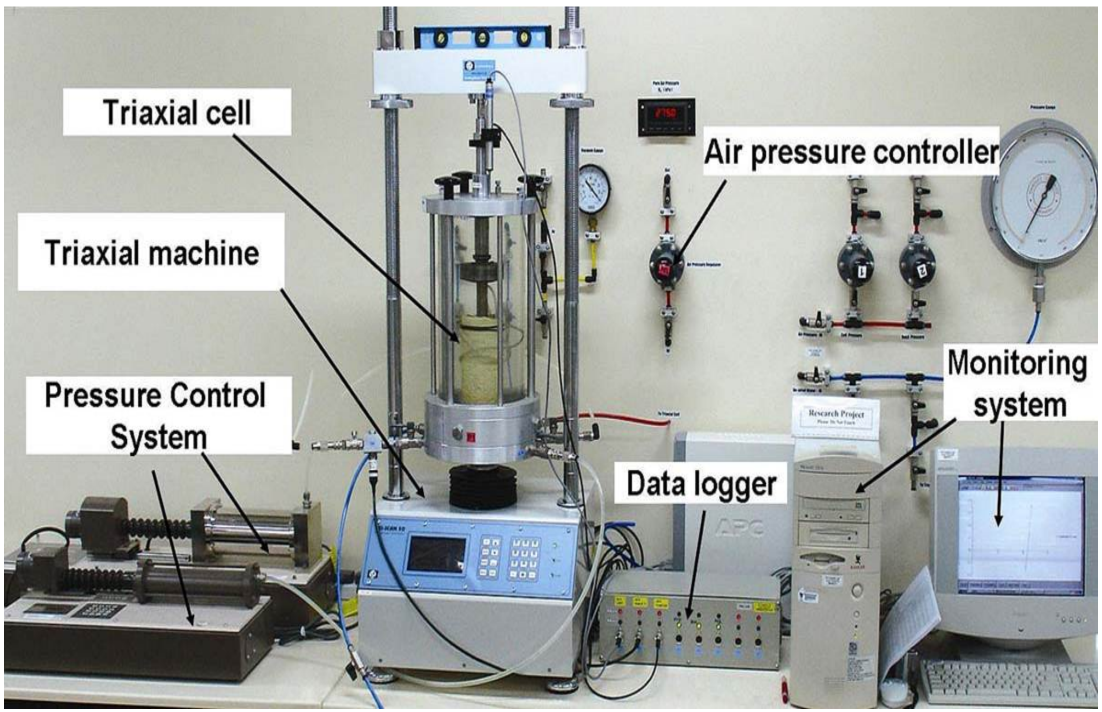
- Face recognition in social media
- Recommendations in online shopping
- Web search: find 1-10 most relevant pages out of millions
- Landslide prediction
- Precision Farming
- Design of structures
- Climate change
- Design of drug molecules, …

TAG – IIT Mandi

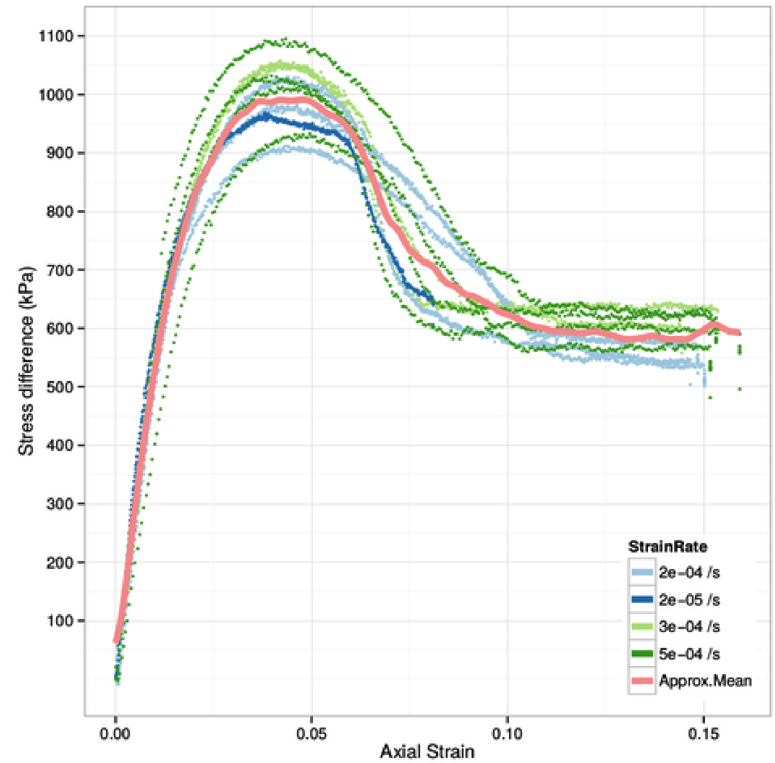
Landslides



STRONG
Soil Strength
WEAK
Soil Moisture Content
DRY
WET



Triaxial cell
Air pressure controller
Triaxial machine
Monitoring system
Pressure Control System
Data logger



Stress difference (kPa)
1100
1000
900
800
700
600
500
400
300
200
100
0.00
0.05
0.10
0.15
Axial Strain
StrainRate
2e−04 /s
2e−05 /s
3e−04 /s
5e−04 /s
Approx.Mean

… Landslides

- Measure soil strength under varying conditions
  - Soil type
  - Moisture content
  - Containment pressure
- Predict/mitigate landslides
  - Depends on weather
  - Load – trees, vegetation, buildings
  - Stress due to nearby traffic, etc
- Cannot measure all combinations

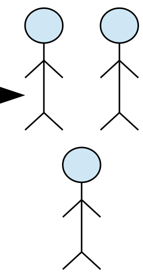
  → interpolate/extrapolate using Data Science

Agriculture

Traditional knowledge

globalisation, climate change

Market



Farm

Produce

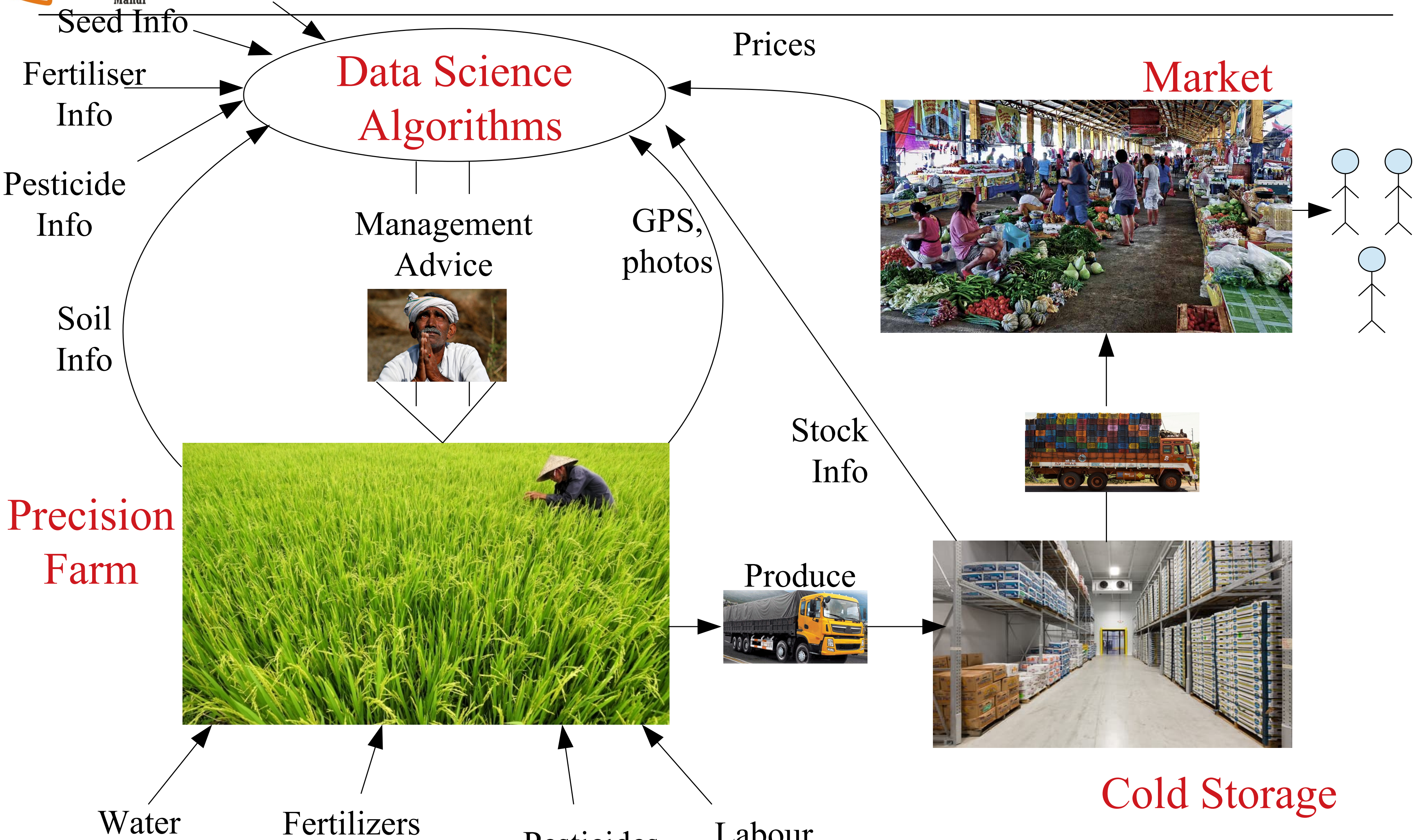
Water

Fertilizers

Pesticides

Labour

Precision Agriculture



… Precision Agriculture

Goal: a decision support system for whole farm management to optimise returns on inputs with minimal use of resources using Data Science

- Relate productivity with geography, environment, water, fertilizer
- Early warning of stresses and disease
- Determine ideal crop variety specific to the locality
- Determine best practices including crop rotation

Input info: weather, market prices, seed varieties, pesticides and fertilisers, soil conditions, ...

FarmerZone™ @ IIT Mandi *a DBT initiative*

Conclusion

- Fundamental laws of technology and market forces drive evolution of computer hardware
- Technology disruptions and societal needs drive scientific and engineering R&D
- Hardware: Cray → SMP → Clusters → Clouds
- Software: Fortran → Parallel Fortran → C → Python
- Scientific computing: Equations → Patterns
- Computing paradigm: Sequential → Parallel → Distributed → Data Science

Books