Microwatt to Megawatt - Transforming Edge to Data Centre Insights

Steve Langridge

steve.langridge@huawei.com

May 3, 2015

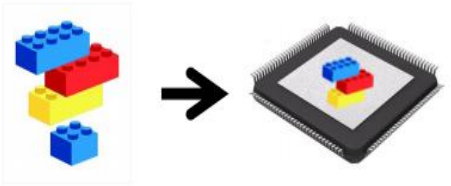
www.huawei.com

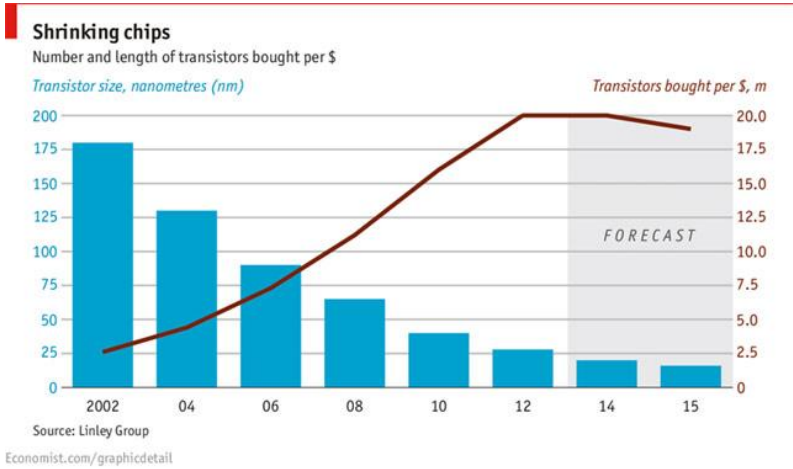
Agenda

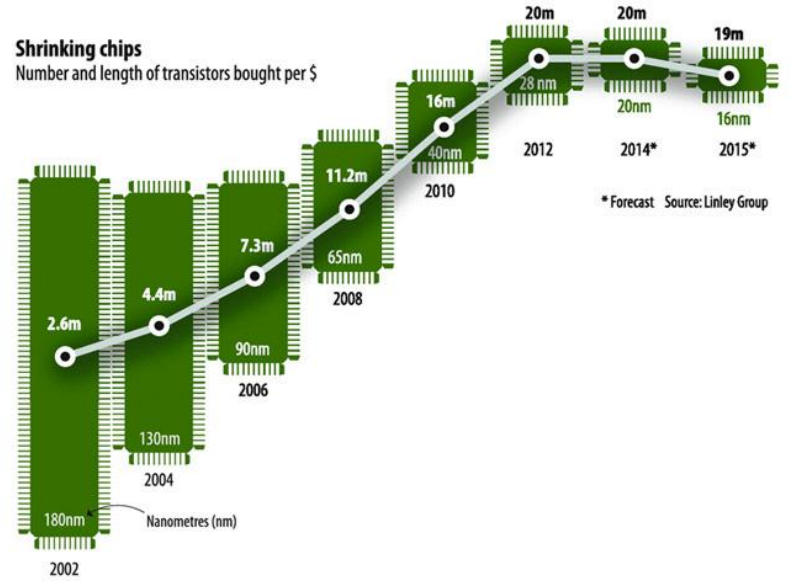
- **HW Acceleration**
- **System thinking**
- **Big Data**
- **Edge to Data Center**
- **Practical examples**

Making a case for Hardware Acceleration

- **The murmur is the end of Moore's Law**
- **Man, this is getting expensive!**
- **The rise of the (vector) machine!**
- **SIMD: A good start**
- **Too hard to put it all together**
- **Industry requires better, faster, cheaper custom silicon**
- **Balancing the new and old workloads**
  - Old: SIMD / Fortran / Vectorizing
  - New: Map-reduce and beyond / Sparse Matrices / Pointer Chasing





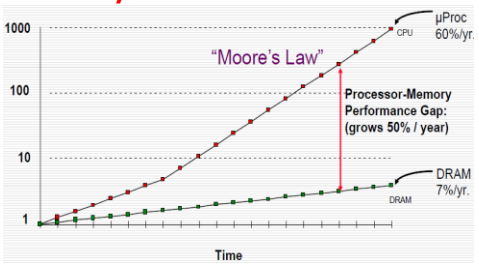


Hardware Acceleration Technology Trends

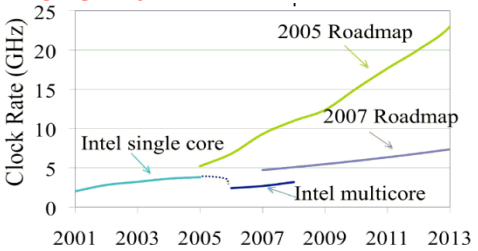
- According to UC Berkeley research result, in the future, General Purpose Processor will meet a Brick wall ;
- Hardware Acceleration will become a better solution to improve system performance;

Brick Wall = Power Wall + Memory Wall + ILP Wall

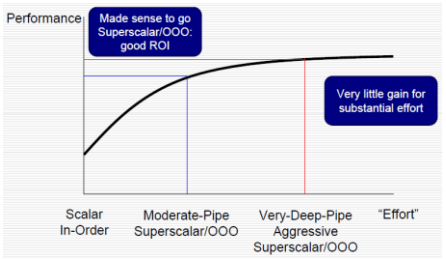
Memory Wall



Power Wall



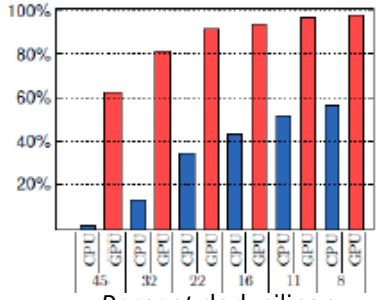
ILP Wall



- Increasing the number of cores increases the demanded memory bandwidth（3D）
- Low-Power Design:
  - Circuit and gate level methods
    - ✓ Voltage scaling, Transistor sizing, Glitch suppression, Pass-transistor logic, Pseudo-nMOS logic, Multi-threshold gates
  - Functional and architectural methods
    - ✓ Clock gating, Clock frequency reduction, Supply voltage reduction, Power down/off, Algorithmic and software techniques,
- ILP Design:
  - Compiler optimization
  - Architecture innovation: branch prediction, Out-of-order execution, Speculation, Very Long Instruction Word

Dark Silicon and the End of Multicore Scaling

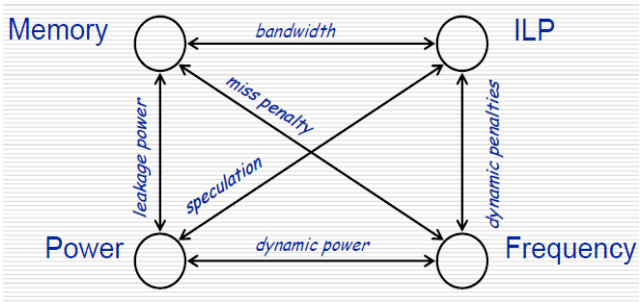
This paper considers all those factors together, projecting upper-bound performance achievable through multicore scaling, and measuring the effects of non-ideal device scaling, including the *percentage of "dark silicon" (transistor under-utilization) on future multicore chips.*



Percent dark silicon

At 22 nm (i.e. in 2012), 21% of the chip will be dark and at 8 nm, over 50% of the chip will not be utilized using ITRS scaling.

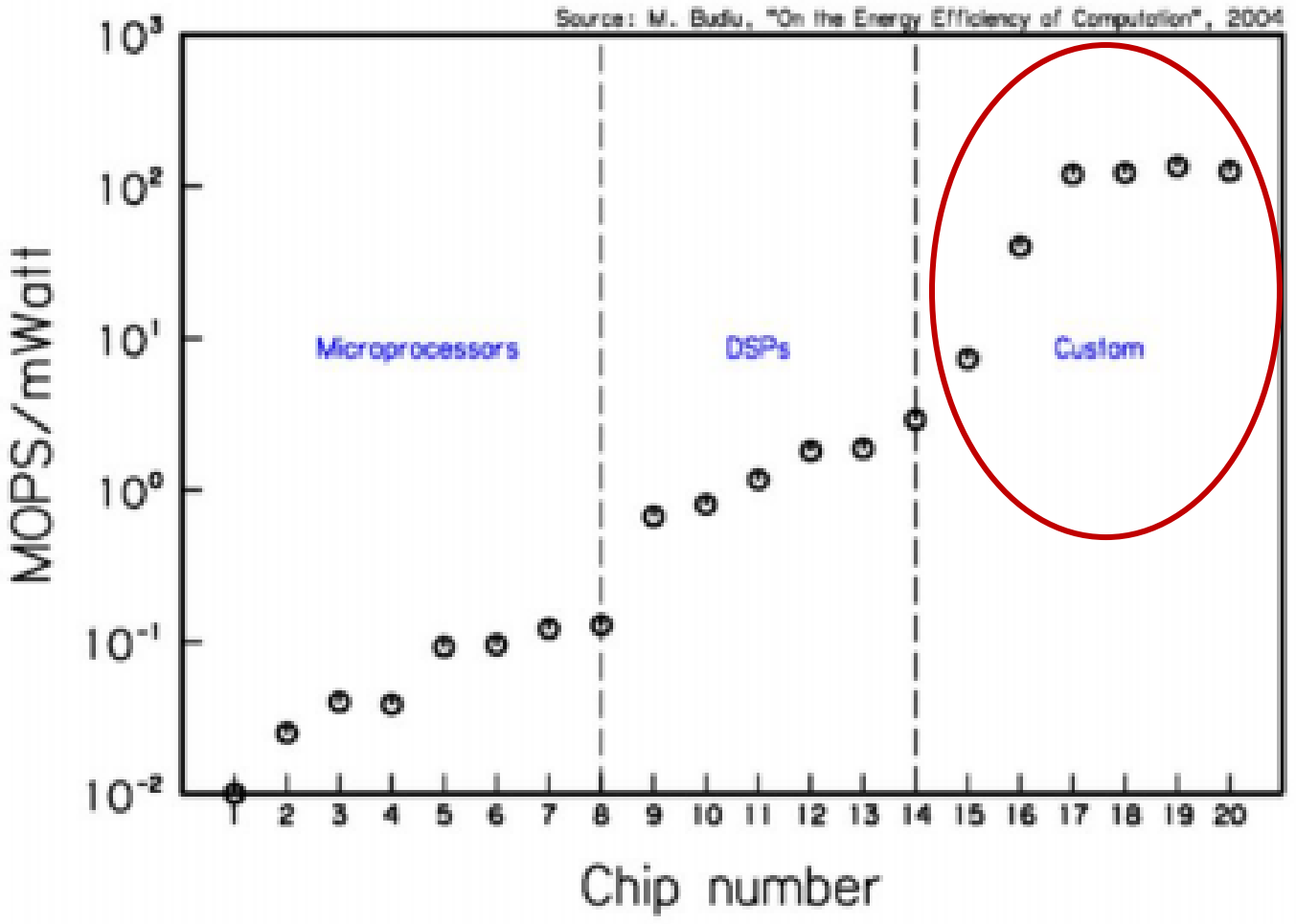
Improving one property comes at the expense of the other



How to improve the performance of system? Hardware Acceleration

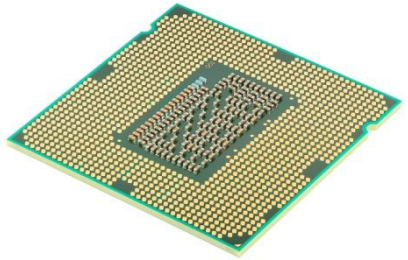
Efficient Processing vs. Complexity

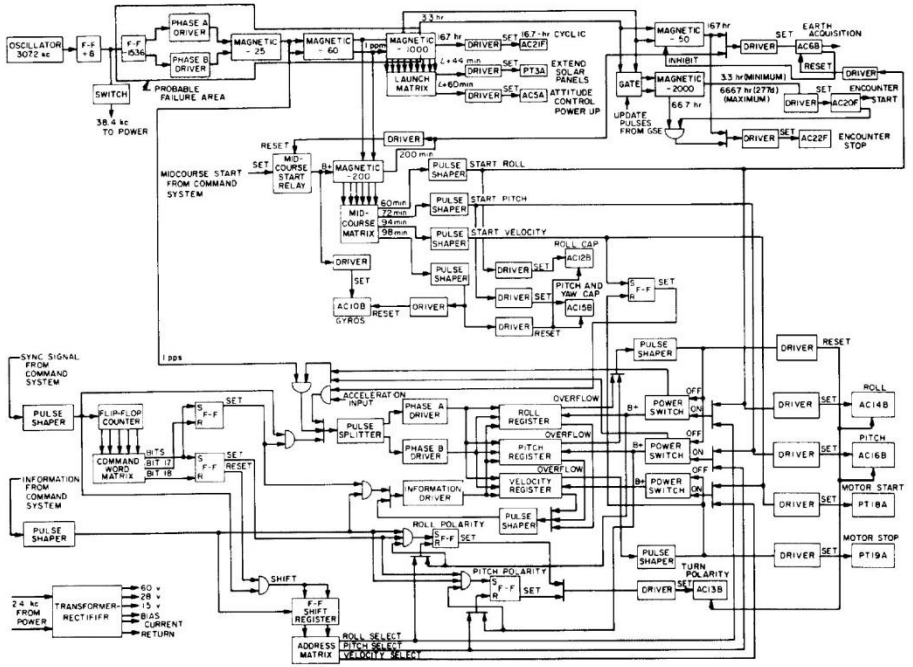
- **Required: the energy efficiency of Custom Processors at the generally adopted programming methods of Microprocessors**
- **[1]Only 15-20% of General Purpose Processors are for the real work of the algorithm**



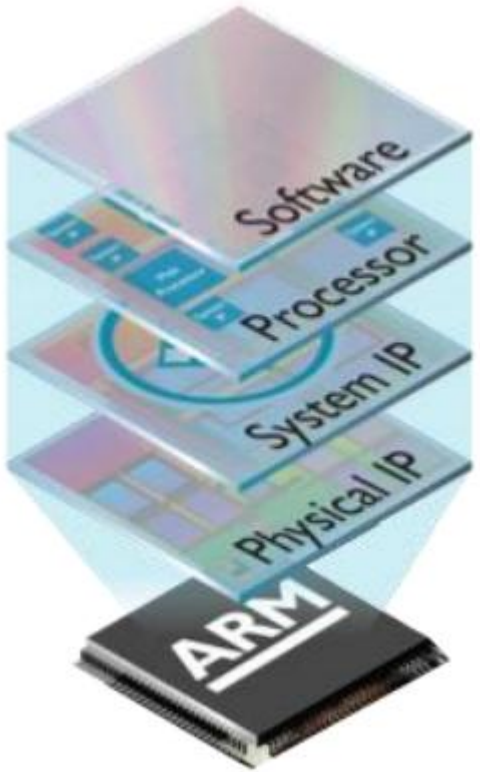
[1] http://www.lanl.gov/conferences/salishan/salishan2013/Astfalk.pdf

Today, solutions are a bit of a ‘parts’ problem





Aha! There are some thoughts about these issues

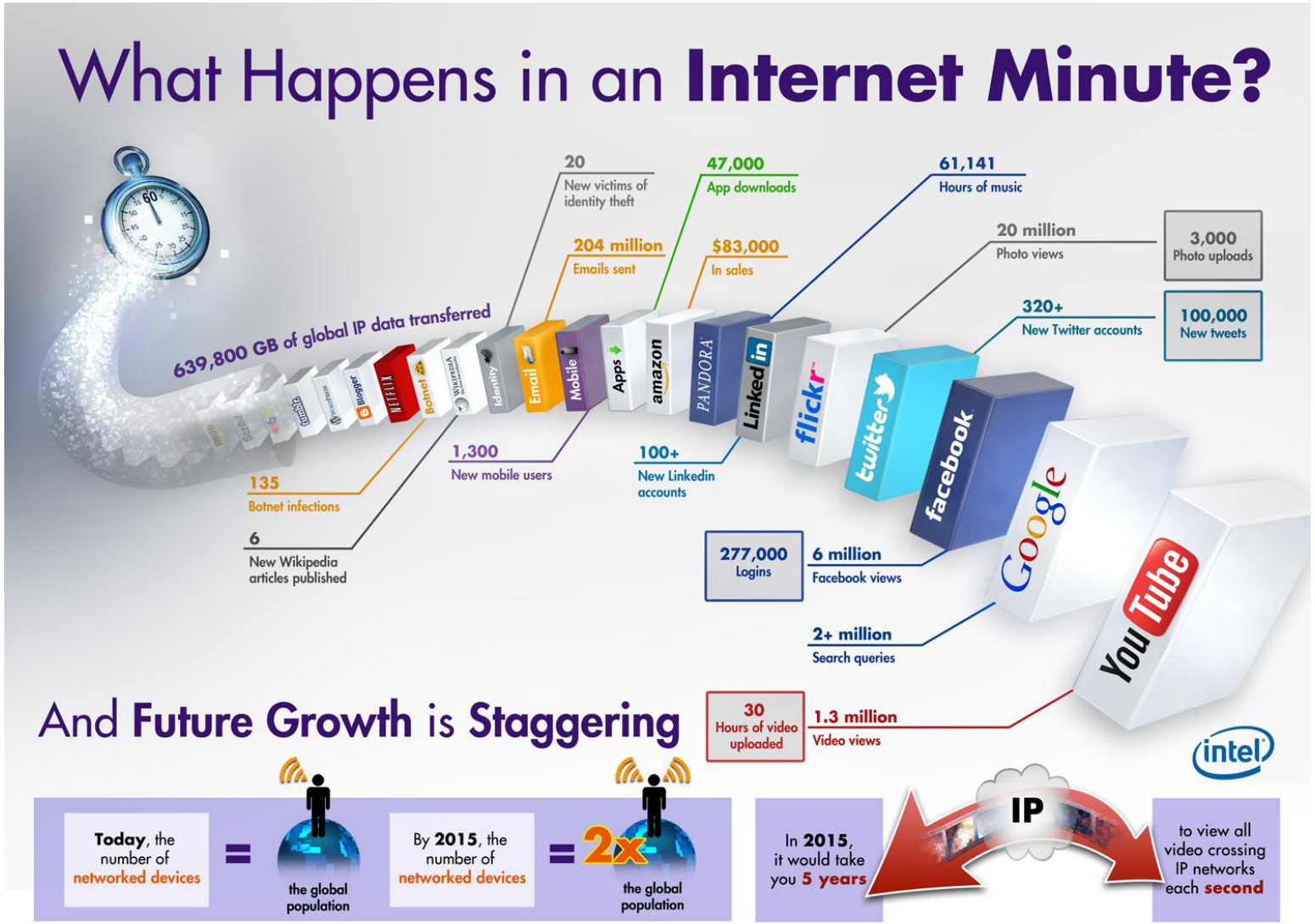


OpenCL

python

Scala

Enter: Big Data



.. And the inevitable sarcasm

Really, Big Data (value) is a Work in Progress

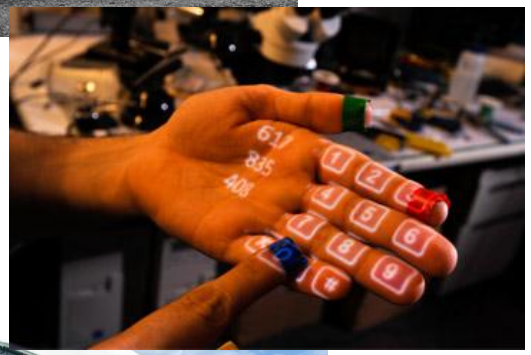
Apache

Spark

STORM

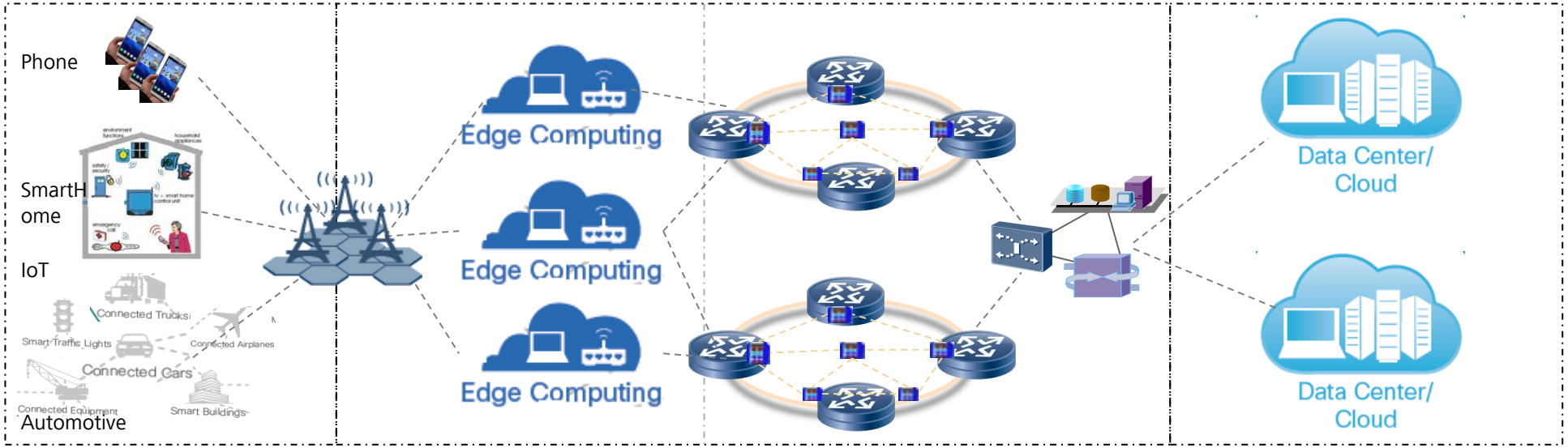
hadoop

We need to re-think **where** we process data, **how** & **why**



20/20 Vision: Micro-watt to Mega-watt

	Device	Edge	Router & Core	Enterprise / Cloud
Operation	Generate / Pre-process	Extract, Transform & Transmit		Load, Process & Analyze
Opportunity	Fast user response	Change “pipe” into iChannels. Processing in the path to Data Center		Processing & Analytics



- ❖ Device
- ❖ Automotive
- ❖ Smart Home
- ❖ IoT

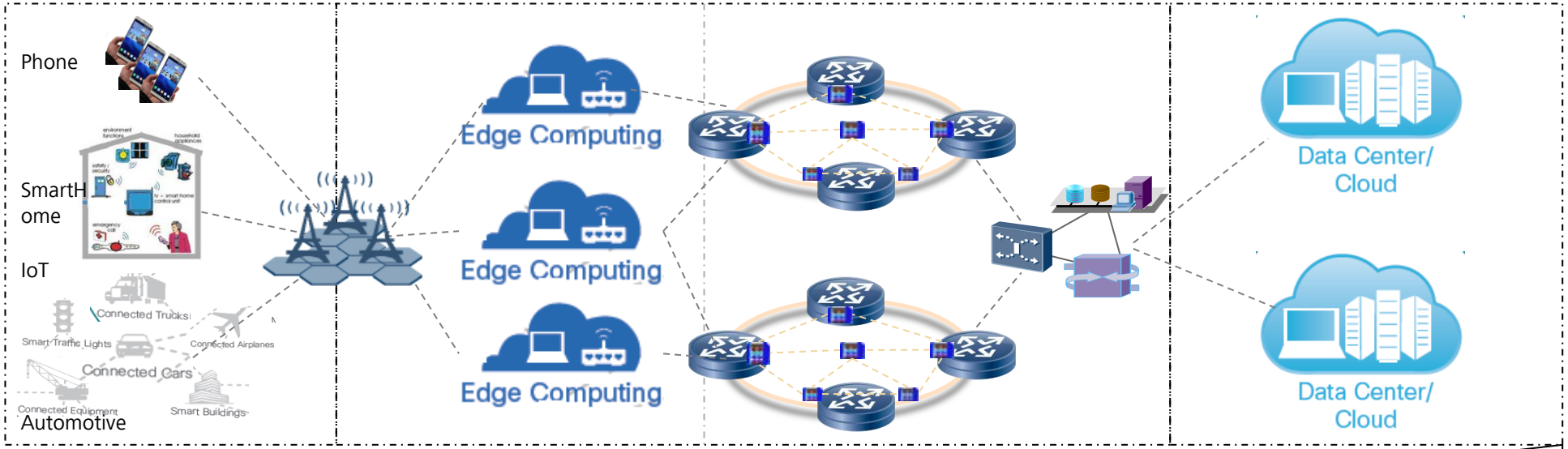
- ❖ Edge

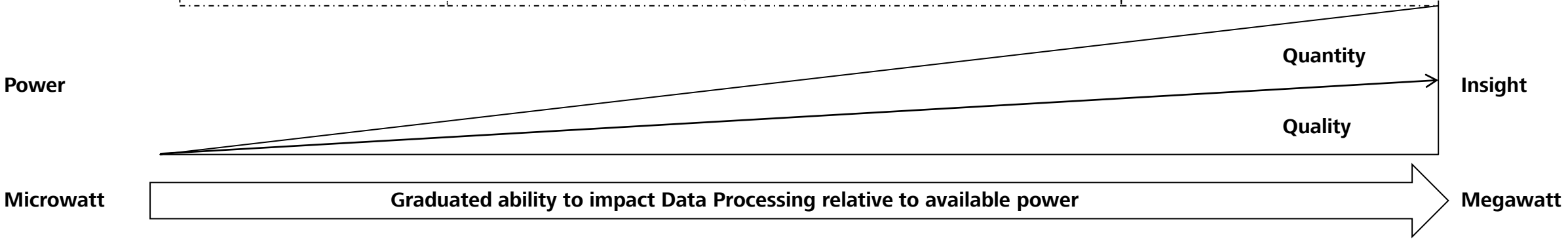
- ❖ Router & Core

- ❖ Cloud Infrastructure

20/20 Vision: Micro-watt to Mega-watt

	Device	Edge	Router & Core	Enterprise / Cloud
Operation	Generate / Pre-process	Transform & Transmit		Load, Process & Analyze
Opportunity	Fast user response	Change “pipe” into iChannels. Processing in the path to Data Center		Processing & Analytics

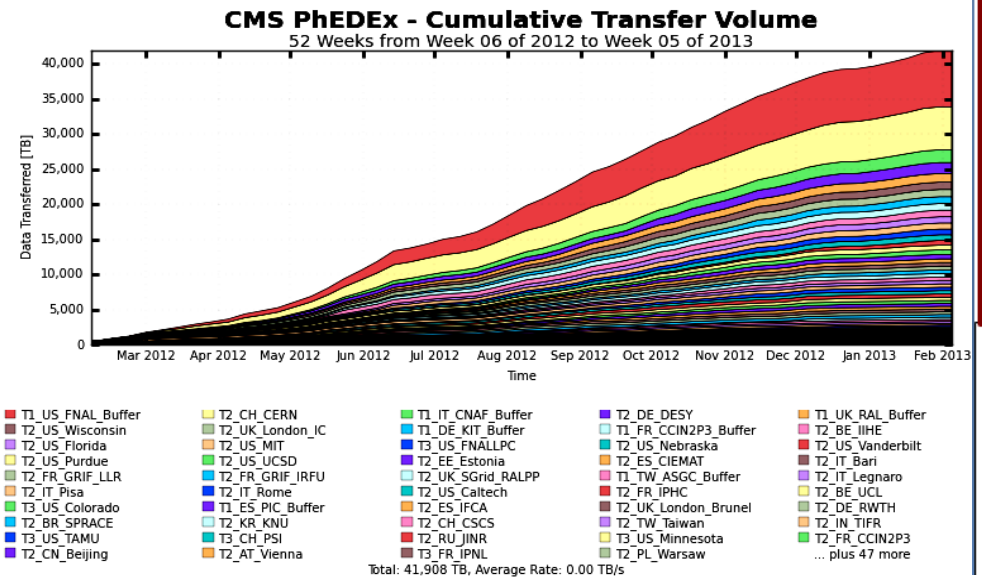


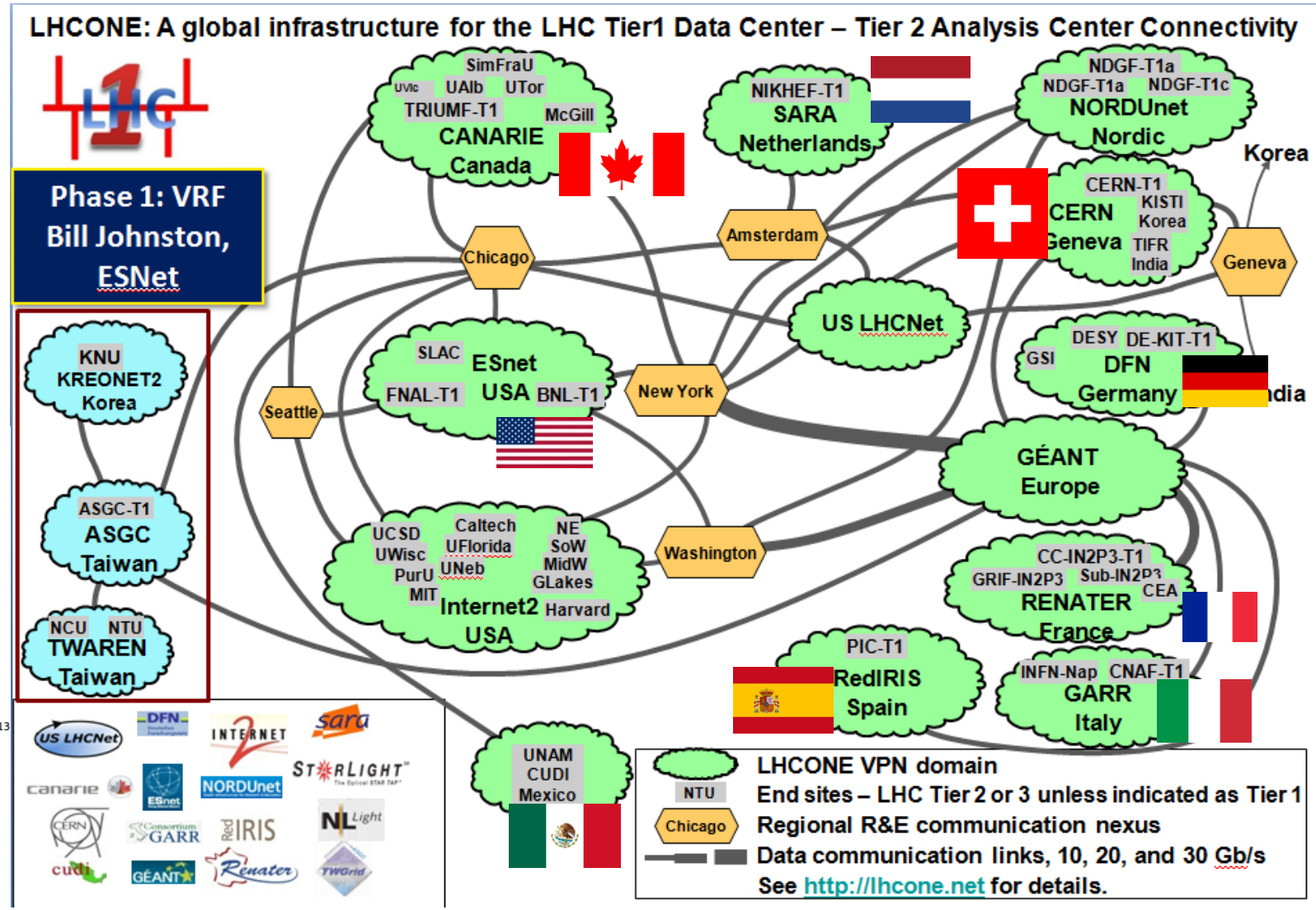


uW>MW Inspiring Example: Large Hadron Collider (CERN)

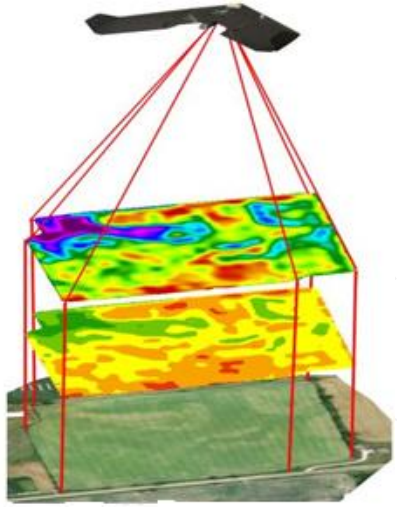
LHC Analysis Network: Producer & Consumer for Big Data analysis
Today: Produce, Store & Distribute raw data to subscribing institutions.

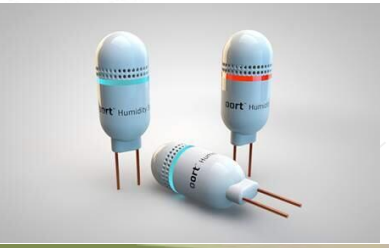
How much (raw) transmitted data is processed with the same algorithms at destination?





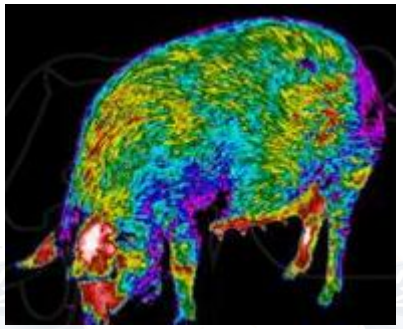
Precision Agriculture – Farm to National Coverage





Farm

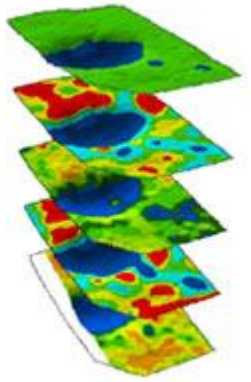
- Secure Wireless
- Drones use CV for Crops and Livestock
- Process data locally and drive automated water & nutrient system efficiency



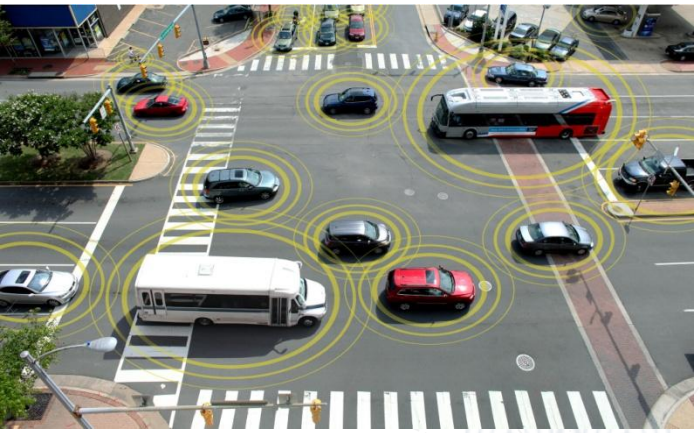
Country

-Food Supply Chain
-Insurance / Risk model
-Early detection of animal-borne disease
-Less / No Crop Subsidy

Aggregation of Regional, Municipal, Provincial and Federal data – mass scale modelling



Traffic Systems



Parking – availablity, rate, time to final destination, services

Congestion re-routing – Without causing knock-on problems!

Weather Impact
Driver Scoring
Fleet Management
Fuel Efficiency

Massively sensored Vehicle Platforms, operating in the context of congested urban setting

Closing thought – It’s just the beginning

Thank you

www.huawei.com