

CO405H

Computing in Space with OpenSPL

Topic 13: 3D Space (Part 1)

Oskar Mencer

Georgi Gaydadjiev

Department of Computing
Imperial College London

<http://www.doc.ic.ac.uk/~oskar/>

<http://www.doc.ic.ac.uk/~georgig/>

CO405H course page:

WebIDE:

OpenSPL consortium page:

<http://cc.doc.ic.ac.uk/openspl14/>

<http://openspl.doc.ic.ac.uk>

<http://www.openspl.org>

o.mencer@imperial.ac.uk

g.gaydadjiev@imperial.ac.uk

Why Computing in Space?


- Reducing costs
- Dealing with insane amounts of data (Big Data)
- Increasing speed (latency and throughput)
- Supporting growth at a very large scale
- Increasing competitive advantage
- Doing something that could not be done before...

- Cloud Computing!!!

“In the cloud, nobody knows that you’re a DFE...”

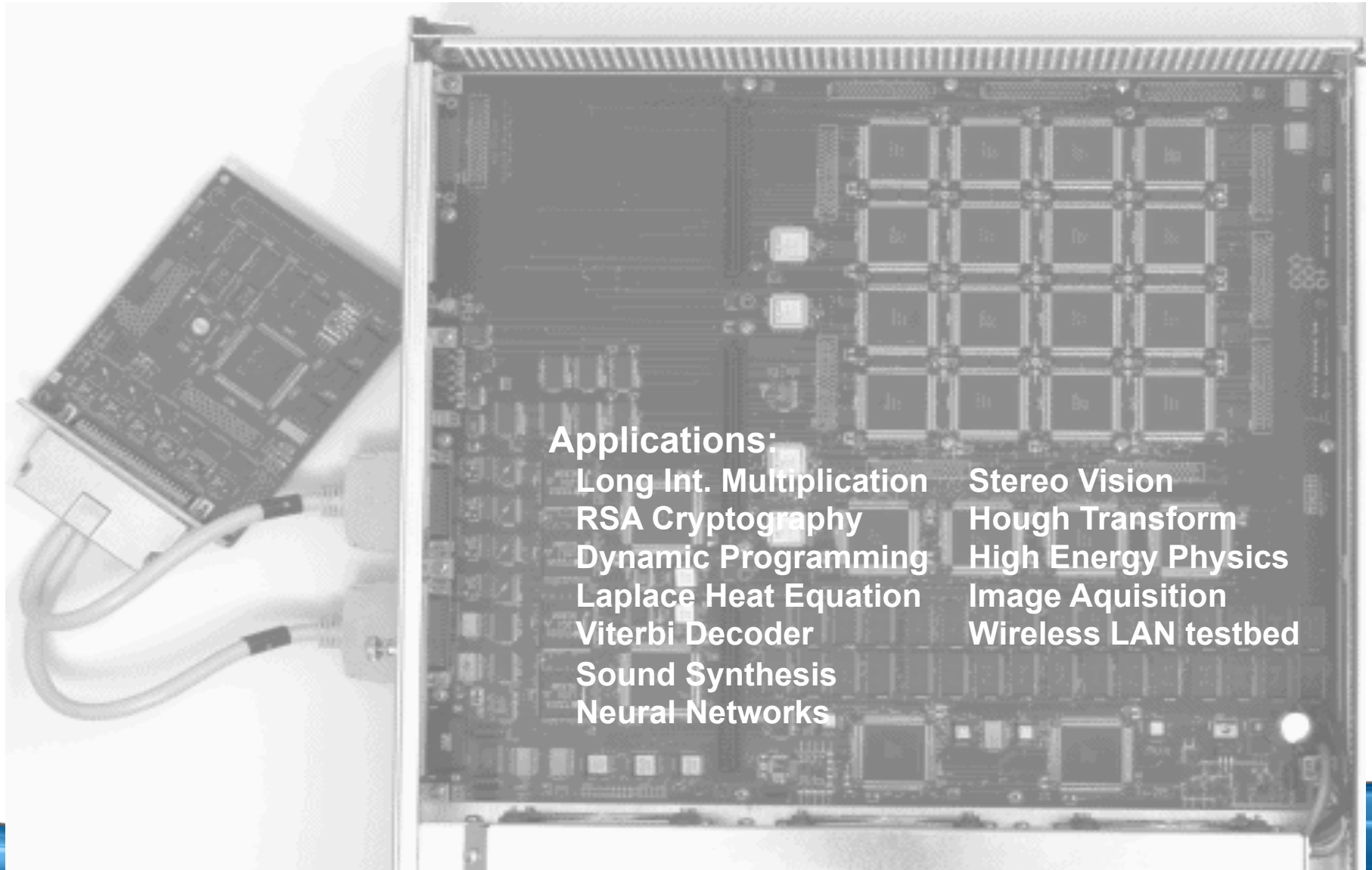


What is Computing in 3D Space

- **First, fix the size of the computer in 3D space**
 - Then program a CPU machine and a Spatial Computer of the same size, and see the gain from computing in space
 - Find a large enough dataset to fill the DFE machine
- 

Computing in Space (1992)

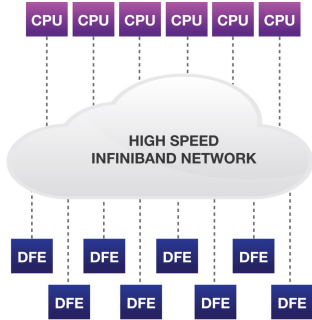
Digital Equipment Corporation DEC PeRLe-1



Applications:

Long Int. Multiplication	Stereo Vision
RSA Cryptography	Hough Transform
Dynamic Programming	High Energy Physics
Laplace Heat Equation	Image Acquisition
Viterbi Decoder	Wireless LAN testbed
Sound Synthesis	
Neural Networks	

MPC-X 2012 Scalable Dataflow Computing



MPC-X Architecture

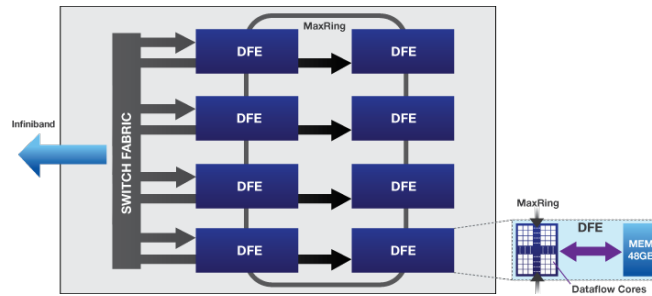
- Combine CPU and DFE nodes to handle unique compute challenges efficiently
- Use low-latency, high bandwidth Infiniband fabric for expandable compute
- Multi-Scale Cluster optimization balances resources at cluster, rack, node and DFE level



1U form “pizza box” form factor

MPC-X Node

- 112GB/s Infiniband Connectivity provides exceptional zero-copy bandwidth and latency
- 768GB DRAM to contain massive datasets
- 2GB/s MaxRing for intra-node communication
- PSU redundancy and ‘lights out’ management



42	
41	Ethernet switch
40	
39	
38	
37	CPU
36	(manager node)
35	
34	MPC-X
33	MPC-X
32	MPC-X
31	MPC-X
30	MPC-X
29	MPC-X
28	MPC-X
27	MPC-X
26	Infiniband Switch
25	CPU
24	CPU
23	CPU
22	CPU
21	CPU
20	CPU
19	CPU
18	CPU
17	MPC-X
16	MPC-X
15	MPC-X
14	MPC-X
13	MPC-X
12	MPC-X
11	MPC-X
10	MPC-X
9	Infiniband Switch
8	CPU
7	CPU
6	CPU
5	CPU
4	CPU
3	CPU
2	CPU
1	CPU

OpenSPL enabled optimizations

Multiple scales of computing

Important features for optimization

complete system level

⇒ balance compute, storage and IO

parallel node level

⇒ maximize utilization of compute and interconnect

microarchitecture level

⇒ minimize data movement

arithmetic level

⇒ tradeoff range, precision and accuracy
= discretize in time, space and value

bit level

⇒ encode and add redundancy

transistor level

⇒ create the illusion of '0' and '1'

And more, e.g., trade Communication (Time) for Computation (Space)

Designed for educational use only using Maxeler Technologies' curve construction methodology. This tool uses delayed data and displayed results are indicative representations only.

Please hover your mouse pointer over column titles and links for further information.

CME Ticker	Bloomberg Ticker	DSF Pricing					Timestamp
		Price	Coupon	PV01	NPV	Implied Rate	
T1UM4 2Y	CTPM4	100'057	0.750%	\$19.97	\$179.69	0.6600%	4:00:03 PM CT 4/4/2014
F1UM4 5Y	CFPM4	100'115	2.000%	\$48.49	\$359.38	1.9259%	4:00:03 PM CT 4/4/2014
N1UM4 10Y	CNPM4	100'225	3.000%	\$90.16	\$703.12	2.9220%	4:00:03 PM CT 4/4/2014
B1UM4 30Y	CBPM4	102'270	3.750%	\$195.07	\$2,843.75	3.6042%	4:00:03 PM CT 4/4/2014
T1UU4 2Y	CTPU4	100'085	1.000%	\$19.93	\$265.62	0.8668%	4:00:03 PM CT 4/4/2014
F1UU4 5Y	CFPU4	100'110	2.250%	\$48.27	\$343.75	2.1788%	4:00:03 PM CT 4/4/2014
N1UU4 10Y	CNPU4	101'125	3.250%	\$89.55	\$1,390.62	3.0948%	4:00:03 PM CT 4/4/2014
B1UU4 30Y	CBPU4	106'020	4.000%	\$193.47	\$6,062.50	3.6868%	4:00:03 PM CT 4/4/2014

Quotes and analytics are updated every 15 minutes.

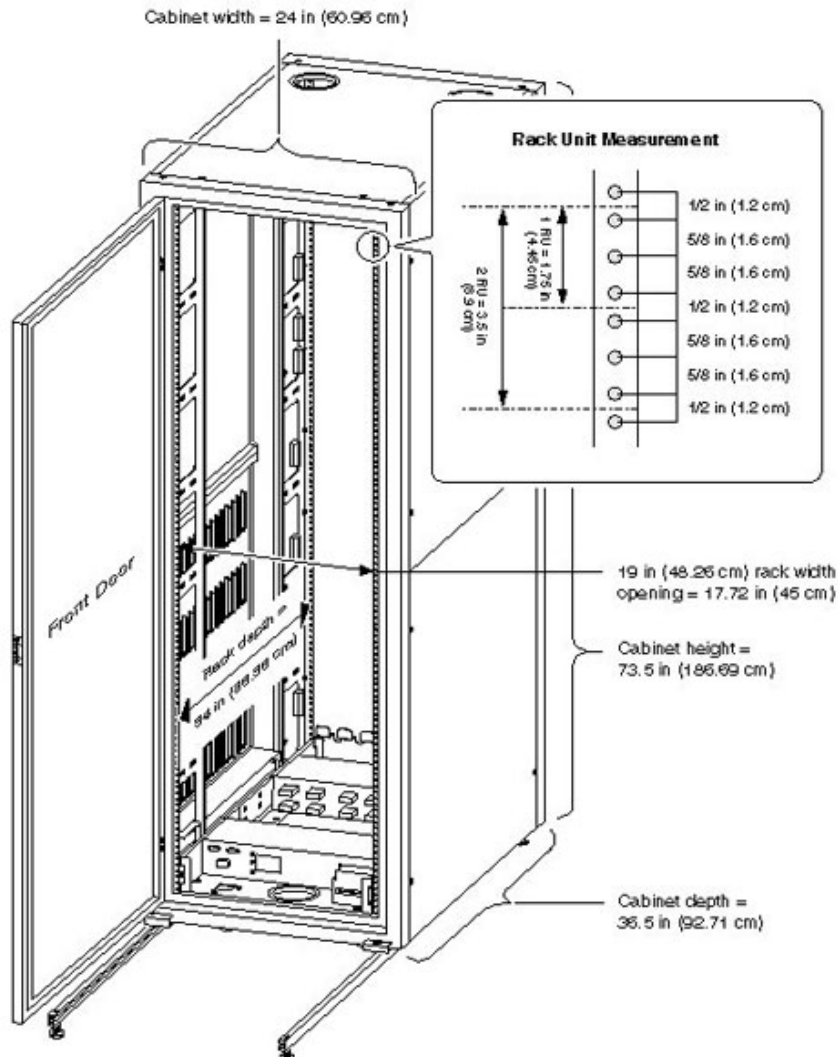
 Analytics powered by Maxeler Technologies®

Instrument	CPU 1U-Node	Max 1U-Node	Comparison
European Swaptions	848,000	35,544,000	42x
American Options	38,400,000	720,000,000	19x
European Options	32,000,000	7,080,000,000	221x
Bermudan Swaptions	296	6,666	23x
Vanilla Swaps	176,000	32,800,000	186x
CDS	432,000	13,904,000	32x
CDS Bootstrap	14,000	872,000	62x

3D Spatial perspective on OPEX: *Measuring Rack Power at the Socket*

- Measurement includes:
 - Cooling, A/C
 - Power supplies
 - Rack level power distribution
 - Networking
 - Storage
 - Memory Chips
 - Compute Chips (CPUs and DFEs)
 - all the other stuff inside a rack that needs electricity
- Power per rack is nice but really we want *Useful Computations per Watt* of rack power consumption

1U definition



A rack holds computing units of 1U, 2U, 3U, 4U

1U = 19inch × 36.5inch × **1.75inch**

2U = 19inch × 36.5inch × **3.5inch**

Each compute unit has it's own power supply

1U CPU servers can have 1-2 mother boards each with multiple CPU chips each with multiple cores.

1U DFE boxes can have 8 DFEs or [6 DFEs + 2 CPU chips], connected via Infiniband

Each 1U DFE box has 384GB of LMEM and 48 independent memory channels.

Spatial perspective on OPEX: *Measuring the 3D Space for a program*

1. Need a good reference point:

Running multithreaded programs on multiple cores, how many cores are there in a 1U box? How many DFEs?

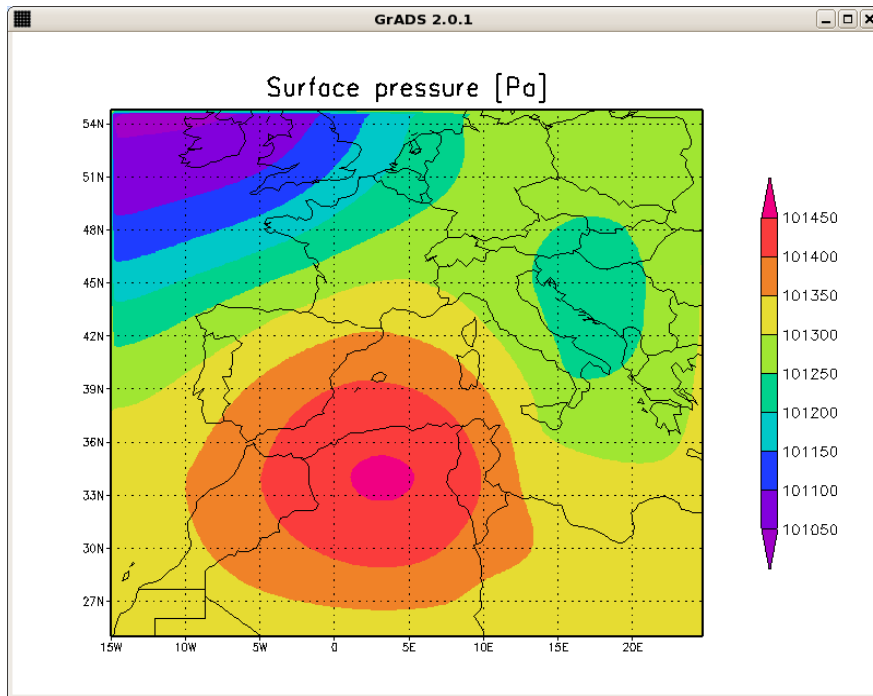
2. How many SysAdmins do you need per rack?

3. Do you measure computations per rack or TeraBytes processed per rack per second?

Evaluate performance by comparing 1U to 1U!

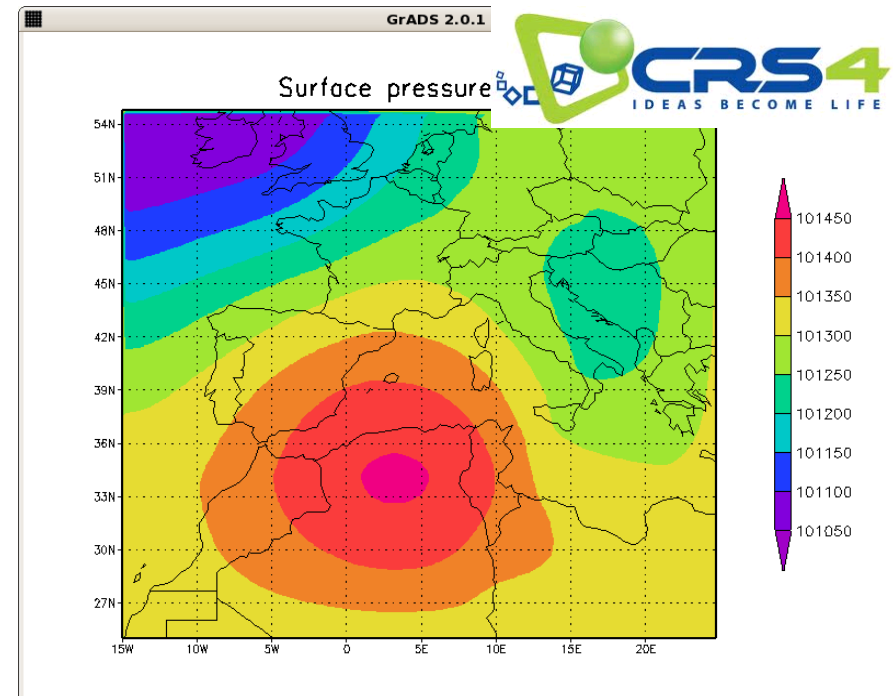


Imaging Platform Example: Weather



1U CPU Node

Wall Clock Time: 2 hours



1U Dataflow Node

less than 2 minutes

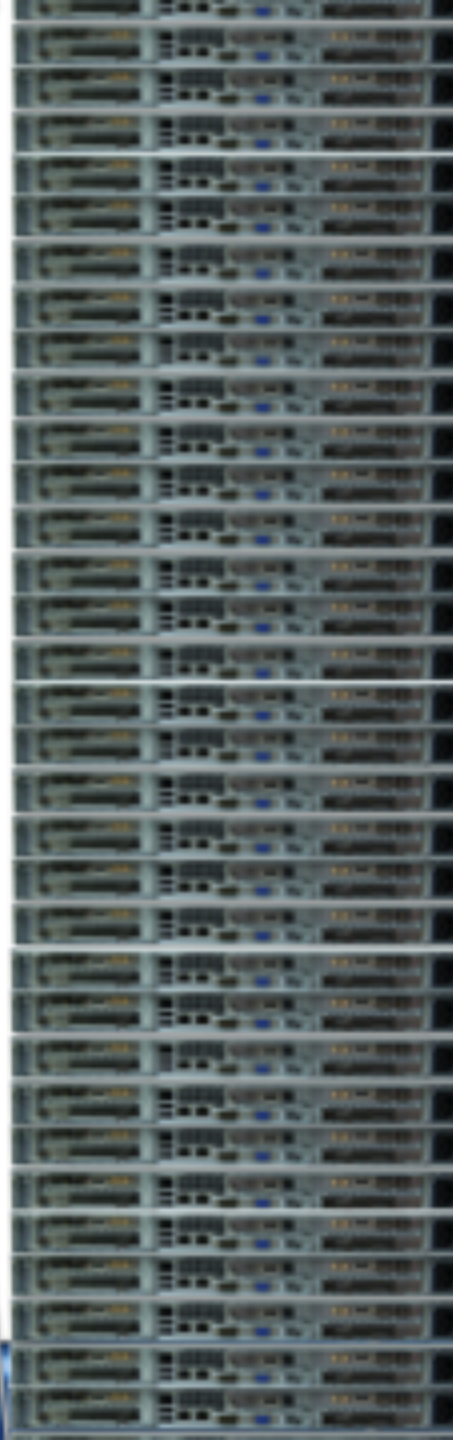
Problem size: (Longitude) 13,600 Km x (Latitude) 3330 Km
Simulation of baroclinic instability after 500 time steps.

Acceleration of a Meteorological Limited Area Model with Dataflow Engines, D. Oriato, S. Tilbury (Maxeler), M. Marrocu, G. Pusceddu (CRS4), SAAHPC Conference, May 2012.

What does 60x mean?



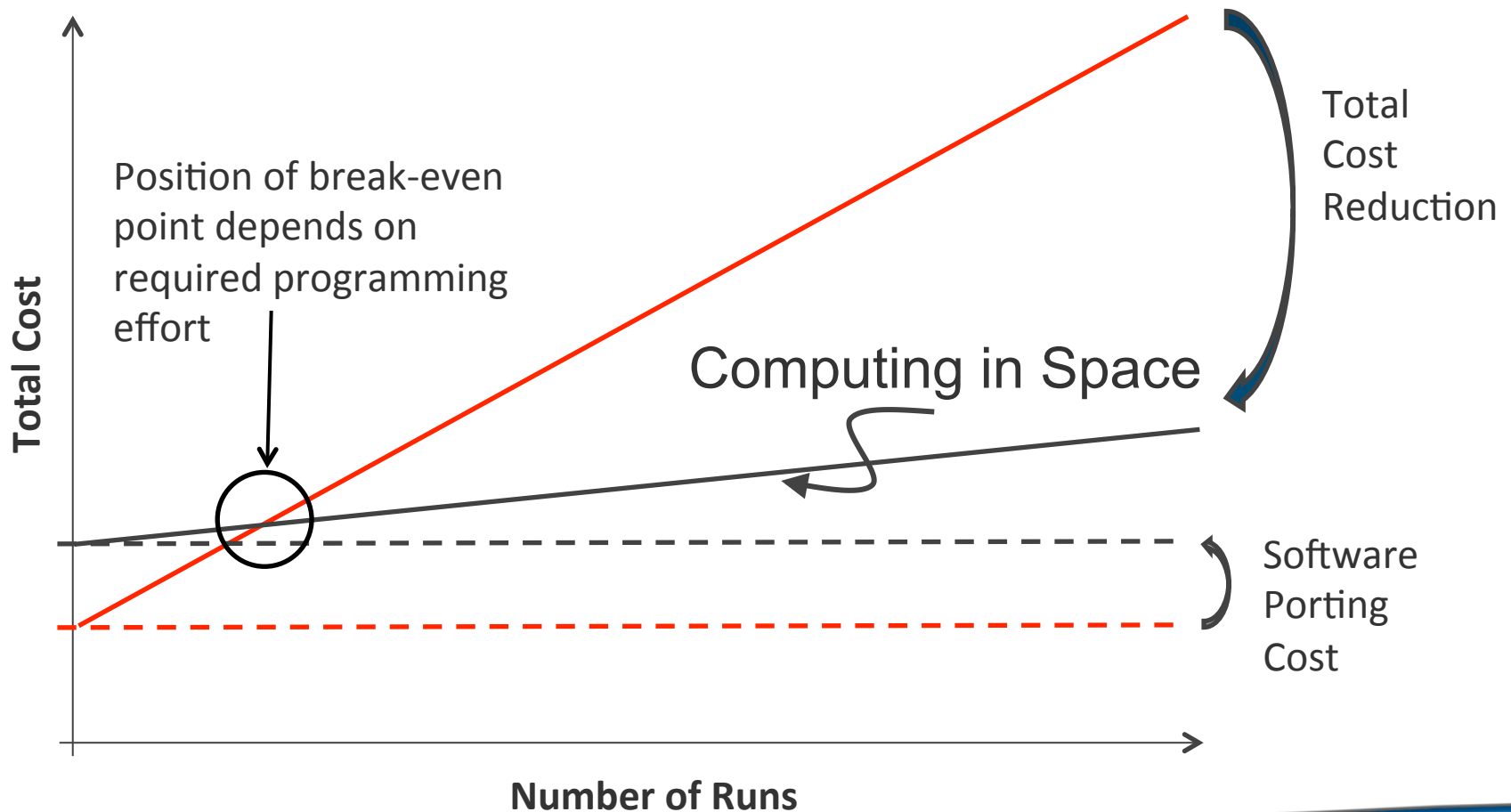
- Putting a supercomputer into a drone with 60x less space/power consumption.
- 6 days => 2.4 hours computing
- 10 images assembled => 600 images assembled
- 1 Mega Watt => 17 Kilo Watts
- Every \$1M spent on buying a dataflow machine delivers power savings over 3 years worth more than \$900K.



Economics of Computing Space

you got to have a lot of data....

Total Cost of Ownership = $f(\text{HW, SW, maintenance, power, real estate, ...})$



Total Cost of Ownership = CAPEX + OPEX

Capital Expenditure + Operating Expenditure

50x

- **50x** Speed-up per 1U server node
- **32** MPC-X Node Solution
- Equivalent to **1600** CPU-only Nodes
- **\$3.2m** Operational cost savings over 3 years

30x

- **30x** Speed-up per 1U server node
- **40** MPC-X Node Solution
- Equivalent to **1200** CPU-only Nodes
- **\$1.8m** Operational cost savings over 3 years

20x

- **20x** Speed-up per 1U server node
- **50** MPC-X Node Solution
- Equivalent to **1000** CPU-only Nodes
- **\$1.7m** Operational cost savings over 3 years

40x

- **40x** Speed-up per 1U server node
- **32** MPC-X Node Solution
- Equivalent to **1280** CPU-only Nodes
- **\$2.6m** Operational cost savings over 3 years

Example of Economics of Computing in Space

On the plus side:

1. Budget for buying a new machine is \$2.8M
2. An x86 CPU-only machine for \$2.8M has \$1.3M annual electricity cost
3. A DFE machine for \$2.8M creates \$27K annual electricity cost

Given that the purchase budget is fixed, the decision to use DFEs saves \$1.27M per year.

AND the DFE machine is a lot faster than the CPU machine with the same purchase price.

So by using DFEs, it is possible to get the advantage of saving operational costs AND at the same time get a machine that is a lot faster.

On the minus side:

Have to port applications to DFEs, and operational savings might not be attractive to everyone...