Babylonian Dynasty
1894-1595 BC

Phil Emma
Chief Scientist
IBM TJ Watson Research
Sept. 2014

In Honor of Yale Patt’s
Dodranscentennial
LXXV

Space: The Final Frontier
1966-1969

The Future of Computer Architecture

3600 Years

The Cell Phone!!!

Thumb’s Up!
Why did God give man opposable thumbs?

Riding the subway?

Playing the guitar?

Hitch-hiking?

Live long, and prosper! (Vulcans)

Comparison of Primate Hands

What can we do because we have thumbs?

Give Up?
Thumbs are for sports...
...to hone our hunting skills.
But we don’t USE our thumbs in most sports, and we generally don’t use them for hunting!

Babylonian Dynasty
1894-1595 BC

Space: The Final Frontier
1966-1969

So what are thumbs for?
For Texting !!!
1) We are too pessimistic in forecasting when many technologies will happen.

2) And we are almost always wrong about how, exactly, they will manifest.

More on thumbs...
Thumbs were first used for counting. That’s why only people and monkeys are known for using Binary.

Aborigines felt no need to count, and while they all had words for “one” and “two,” only some made it to “three.” The Walpiri, for example, only has words for “one,” “two,” and “many.”

The Piraha of the Amazon have also been cited as using a “one-two-many” system of counting.

This is why our first few generations of computers have used the Binary number system.
As we changed to Sexagesimal (Base 60 – used by the Sumerians in the 3rd Millenium BC, and then the Babylonians), Two Thumbs UP, in addition to “Many,” was used as a sign of ultimate approval, or of “Excellence;” also it’s meaning in Decimal (Base 10) today.
I am thinking about something much more important than bombs. I am thinking about computers. - John von Neumann, 1946
Philosophical Breakthroughs

• Babbage – *Automatic calculation by machine*. Manipulating numbers can be done mechanically – without “thought.”

• Turing – *Theory of computation*.

• Eckert & Mauchley – *Doing many fast calculations on a very large scale*.

• von Neumann – *A stored-program computer*. The program can modify itself while running: what the program *does* depends on the data.
How Many View Moore’s Law

$\log$(whatever)

The Good-Old Days

Now

Time

The World Ends
Moore’s Law – My Interpretation

CMOS Roadmap: No real architectural innovation required. Keep scaling the Technology!

The Birth of Architectural Innovation !!!

HAPPY BIRTHDAY YALE
Where we’ve been since von Neumann (1946)

Programming
- Binary
- Assembly
- HLL
- Proc.-Oriented
- Object-Oriented

Applications
- Calculations
- Linear Systems
- Text Processing
- OS
- Database
- Transactions

Architecture
- Registers
- Stacks
- Virtual Machines
- VLIW

What dimensions have failed to evolve?
A, B, and C are NOT visible to the program (to the ISA). They can be defined as best fits any specific implementation.

Legacy code – bound to the ISA.
Target Address (following misprediction)

Patt Sub-engine $P_{31}$

Patt Sub-engine $P_{32}$

Patt Sub-engine $P_{33}$

3 I-Fetch Predictions

$P_{31}$ & $P_{33}$ Inputs

$P_{32}$ & $P_{33}$ Inputs

$P_{31}$ & $P_{33}$ Inputs

$P_{31}$ & $P_{32}$ Inputs
Three Big Ones

• Environmental
  – What’s the temperature? (Regulate)
  – How much energy & power? (Regulate)
  – What’s running with me? (Optimizing)
  – Cost of operation? (Dynamic Adjusting)

• Moving Beyond Scalar Operands
  – Vectors, Matrices
  – Records, Linked Lists
  – Objects

• Abstract Computing, Acceleration
  – Direct manipulation of surfaces & objects
  – Sensory operation
    • Hardness & shape
    • Odor & Taste
  – Estimation & approximate computing
  – The Brain???

Duh!
The NEW Definition of the von Neumann Bottleneck!

To the program, storage looks like this. Why? Because different levels of storage know nothing about the physical layouts of other levels:

- **L1**: Read Request
- **16 Bytes**
- **L2**: 8 Bytes, 1/2 Speed

**Why does the OS juggle 100s of jobs today?**

A physical 3D system like this would *simplify* the OS dramatically.

**DDR Memory Cards**

**L3**

**PROC**
Matrix Multiplication in 3D

For Row = 1 to n do
Begin
  For Col = 1 to n do
  Begin
    C[Row,Col] := A[Row] X B[Col];
  End;
End;

If I have multiple planes of data (in 3D), I can deal with multidimensional data directly.

This is a NEW Machine-level Instruction!

Only 2 misses in 3D!

The Cache understands “Row” vs “Column”

All Scalar Code is Gone!

Only 2 Registers Required:
They’re purely “Structural,” i.e., it’s a 2D structure.

This is >> Order of Magnitude Speedup, and uses FEWER registers

If I have multiple planes of data (in 3D), I can deal with multidimensional data directly.
Why doesn’t the ISA or the cache “understand” these structures?
Topography (looking down) One Layer of this Slice Stored as a Binary Image on the 2D Plane Corresponding to a Fixed Altitude 1s everywhere else ???

Manipulation of Topologies

Facial Recognition

Topographical Maps
Power vs. Frequency

3D Core-on-Core Study

Energy Efficient Operation
High Speed Operation
Slope = 3.5

Frequency (Normalized to 4 GHz)

Normalized Power

Energy Efficient Operation
High Speed Operation
Slope = 1.1

50%

25%

50% 25%

Energy Efficient Operation
High Speed Operation
Slope = 1.1

3D Core-on-Core Study

Energy Efficient Operation
High Speed Operation
Slope = 3.5

SPEC
AK6
FNE

98%
98%
95%

CJR CaD
DLM

1 Proc. @ 3.96 GHz with an 8M L3
2 Proc. @ 3 GHz each with a 4M L3
2 Proc. @ 3 GHz with Shared 8M L3
What is “Architecture”? 

The purpose of “Architecture” is to create a good interface between applications and technology – with a classical understanding of applications, and a contemporary grasp of the technology.

Both evolve. And each drives the other.

Architecture requires broad technical knowledge, coupled to a lively imagination, and what I’ll call an “Artistic Sense.” You can calculate parts of what you need, but sheer “facts” are insufficient.

Success: Is it a Blessing or a Curse?
**Academia**

**What Kind of Research Should we Do?**

“A great building must begin with the immeasurable, must go through measurable means when being designed, but in the end must be unmeasured.” – Louis Kahn

“Architecture should have little to do with problem solving - rather it should create desirable conditions and opportunities hitherto thought impossible.” – Cedric Price

“If you want to get rich from writing, then write the sort of thing that’s read by persons who move their lips when reading to themselves.” – Don Marquis

“I want a man who is kind and understanding. Is that too much to ask for in a millionaire?” – Zsa Zsa Gabor