4

## Language: Fragmentation of Machine Architecture

Sean Parent | Sr. Principal Scientist Adobe Software Technology Lab Mario Wolczko | Oracle Labs

## Desktop Applications - Recent History

- Macintosh
- 68K, single-core
- PPC, single-core
- Intel, multi-core, SIMD, OpenGL/CL
- Windows
- Intel, single-core
- Intel, multi-core, SIMD, OpenGL/CL


## Two Key Events

- 2005 we hit the physical limits of Moore's Law under current technology
- 2007 the iPhone is introduced


## 2009 Projected Processor Characteristics



The future of computing beyond Moore's Law, Volume: 37t, Issue: 2166, DOL (10.1098irsta.2019.0061)

## How are we doing?



Desktop Compute Power (8-core 3.5GHz Sandy Bridge + AMD Radeon 6950)


## Platform Expansion

- Mobile
- iPhone fundamentally changed mobile devices
- Web
- Content Ubiquity is expected
- Tablets
- Larger "phones" succeeded where smaller desktops failed


## Platform Expansion

- In 2012 I gave an internal presentation at Adobe on content ubiquity
- Broadband was available to th majority of the popultion in the developed countries
- Soon will be true worldwide
- Noted capabilities of mobile devices
- Increased by > 8x
- Because content ubiquity is becoming a base expectation, not providing it will kill a product


## Platform Expansion



## Platform Fragmentation

- macOS
- Win32 \& UXP
- iOS and iPadOS
- Android
- Linux (server)
- W3C


## Instruction Set Fragmentation

- Intel (AVX SIMD)
- ARM (Neon SIMD)
- WASM (WASM SIMD)
- Currently 32 bit address space


## GPU Platform Fragmentation

- Metal (Apple)
- DX12 (Microsoft)
- Vulcan (Open Standard, Android, Linux)
- CUDA (NVIDIA)
- WebGPU (Browsers)


## Amdahl's Law



## Hardware to Fight Amdahl's Law

- NUMA
- DMA to discrete GPU
- Unified Memory (Apple's M1 chips)
. "The unified memory requires a very different approach to that on Windows with discrete GPUs."
- Optane?


## Hardware to Manage Power

- Thermal Throttling
- Heterogeneous Cores
- Discrete / Integrated GPU Switching


## Languages Are Not Keeping Up

- We are struggling to find models to reason about concurrent systems
- CSP, Actors, Functional, ...
- Safer languages have higher overhead
- But unsafe languages are harder to get correct
- My estimate is we are leaving $2^{3}$ to $2^{5}$ times performance on the table


## Operation Costs are Not Reflected In Code



## What is wrong with C++

- C++ allow us to control memory layout and sharing
- Compiler is blind to sharing - aliasing + mutation kills optimization
- Developer is also blind to sharing making code difficult to reason about
- Lack of safety makes it very difficult especially in the presence of concurrency for new developers
- Basic library primitives for concurrency (threads) are very expensive
- Performance penalty (Stepanov Abstraction Penalty) to wrap basic arithmetic types
- i.e. treat a uint8_t as a value from 0.0-1.0
- Code the same algorithm with different function names or pay the tax

C++

- Despite the limitations and drawbacks, C++ is still performance king*
- Adobe has a massive investment in C++ code bases
- Can C++ be improved enough?
- The pace of $\mathrm{C}++$ advancement is still rapid
- If another language proved to be better, what does the migration look like?


## Possible Future?

- In 2007 I gave a Google TechTalk, A Possible Future of Software Development
- Observation - Most developers cannot write a correct binary search (still true)
- An argument for developing generic libraries and concepts
- Conjecture - All problems of scale become a network problem
- An argument for developing declarative systems
- BNF, SQL, HTML, Spreadsheets
- In imperative languages a single relationship becomes multiple functions


## Imperative Solution to Mini-Image Size



## Declarative Solution using the Property Model Library

```
sheet mini_image_size
{
input:
    original_width : 5* 300;
        original_height : 7 * 300;
interface:
    constrain :true;
    width_pixels : original_width <== round(width_pixels);
    height_pixels : original_height <== round(height_pixels);
    width_percent;
    height_percent;
logic:
    relate {
        width_pixels <== round(width_percent * original_width / 100);
        width_percent <== width_pixels * 100/original_width;
    }
    relate {
        height_pixels <== round(height_percent * original_height / 100);
        height_percent <== height_pixels * 100 / original_height;
    }
    when (constrain) relate {
        width_percent <== height_percent;
        height_percent <== width_percent
        }
output:
    result <== { height: height_pixels, width: width_pixels }
}
```


## Where do programming languages need to go

- Major shift from developer productivity to code efficiency
- Locality, locality, locality
- Data oriented, array based
- Value semantics with safe mutability
- Reference semantics and garbage collectors are problematic
- Computation kernels
- Supporting SIMD and GPU code generation
- See Halide language


## Example of Halide

Func blur_3x3(Func input) \{
Func blur_x, blur_y;
Var x, y, xi, yi;
// The algorithm - no storage or order
blur_x(x, y) = (input(x-1, y) + input(x, y) + input(x+1, y))/3;
blur_y $(x, y)=\left(b l u r_{-} x(x, y-1)+b l u r_{-} x(x, y)+b l u r_{-} x(x, y+1)\right) / 3 ;$
// The schedule - defines order, locality; implies storage
blur_y.tile(x, y, xi, yi, 256, 32)
.vectorize(xi, 8).parallel(y);
blur_x.compute_at(blur_y, x).vectorize(x, 8);
return blur_y;
\}

## Possible?

$$
\begin{aligned}
\text { blur_3x3 }=[2]=> & \left.\left(\left(\left[\begin{array}{cc}
{[1,0]+[0,0}
\end{array}\right]+[1,0]\right) / 3\right) \right\rvert\, \\
& (([0,-1]+[0,0]+[0,1]) / 3) ;
\end{aligned}
$$

## Where do programming languages need to go

- Switch emphasis from safety to correctness
- Higher level semantics allows for more optimization
- Graph based
- Ability to control flow between software components
- Shift from functions to relationships


## Machine Learning - the wild card

- CoreML (Apple)
- DirectML (Microsoft)
- Neural Engine (Apple)
- TPU (Google)


## About the artist

## UV Zhu

With an eye for the abstract, Chinese artist UV Zhu remixes patterns, textures, and colors to explore the future of fashion. Using Adobe Photoshop, Adobe Illustrator, and Maxon Cinema 4D, he blends surreal settings, organic shapes, and even favorite foods to challenge convention. Inspired by his travels-around the Internet and in real life-for this piece, UV fantasized about characters moving through an imaginary world, the things they might do, and what they might wear. The result is a bright, colorful expression of joy and positivity.


