

From Matlab to FPGA in manageable steps, a true story in double precision

Mike Looijmans & Dirk van den Heuvel Topic Embedded Systems







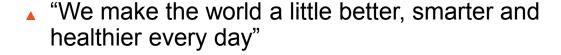


Who are we?









- Real Embedded company; 175 employees
 - 125+ embedded software developers
 - 25+ FPGA designers
 - 10+ board designers
- Founded in 1996, privately owned
- Based in the Netherlands; Europe
- 4 Business Units:
 - Consultancy: The Netherlands
 - Turn Key Projects: Europe and North America
 - Embedded Product Development and Sales: Worldwide
 - Healthcare Solutions: Worldwide















Delirium measurement



- Mission: safe and accurate delirium monitoring in routine hospital care (delirium "thermometer")
- ▲ Delirium (acute brain failure) affects over 3 million hospitalized patients in Europe every year
- ▲ It is a potentially fatal medical emergency that regularly leads to
 - long-term cognitive impairment (dementia)
 - longer hospital admission
 - ▲ higher healthcare costs
- ▲ Delirium equals diabetes in social costs, ~ €173 billion
- The longer a delirium episode lasts, the more damage is done.
- ▲ To date, delirium is detected too little and too late
 - ▲ Subjective and ineffective methods
- http://www.prolira.com
- https://www.youtube.com/watch?v=m8HQImVjHTM





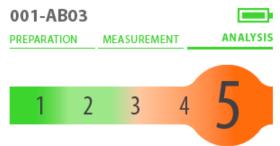




Prolira's DeltaScan patch and monitor

- ▲ EEG-based electrode disposable patch
- ▲ Hardware/software brain activity analyzer
- ▲ Algorithmic recognition of delirium and treatment
 - Patented
 - Originally Matlab modelled
 - ▲ Transition to, translated to Octave
 - ▲ STEP 1: Manual translation to C++
 - ▲ Streaming data processing model
 - ▲ More performance enhancement control





new measurement



What did Topic do?



- ▲ Designed the electronics
- Miami System-on-Modules as core processing platform
- ▲ Developed a V1 and V2 prototype
 - ▲ Sensor cable interface
 - ▲ Display and touch interface
 - ▲ Battery interface
- ▲ Linux based BSP configuration and driver development

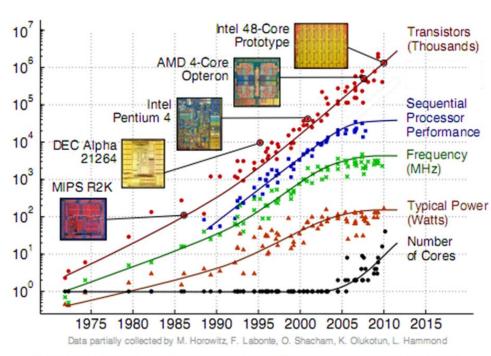
- ▲ Application development support
- ▲ FPGA accelerator design to get the application real-time performing
- ▲ Helped the customer with CE approval and device certification





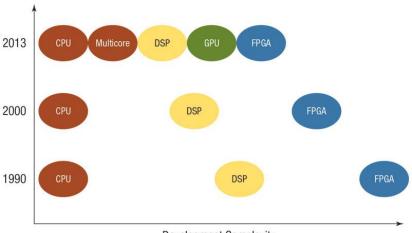
No More Moore's law





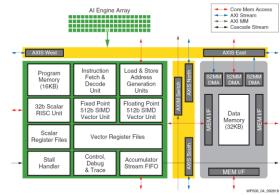
Prepared by C. Batten - School of Electrical and Computer Engineering - Cornell University - 2005 - retrieved Dec 12 2012 - http://www.csl.cornell.edu/courses/ece5950/handouts/ece5950-overview.pdf

http://www.gotw.ca/publications/concurrency-ddj.htm



Development Complexity









Execution acceleration: Wavelet transform

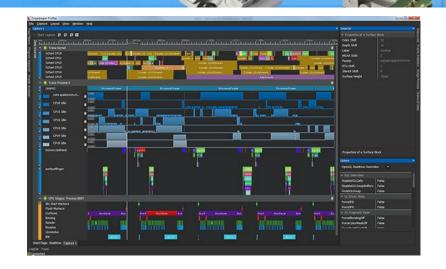
- ▲ Profiling: determine performance bottleneck
 - ▲ Wavelet transform (WT) and inverse (iWT)
 - ▲ 2 msec processing window allowed
 - ▲ 1 wavelet transform takes ~3ms

▲ STEP 2 : Evaluate acceleration options

- ▲ Double precision floating point to fixed point = algorithm stability issues
- ▲ NEON/FPU engine = not sufficient acceleration
- ▲ FPGA = double precision floating point???

▲ Implementation:

- ▲ C function isolation
- ▲ Code optimization
- ▲ Vivado HLS implementation







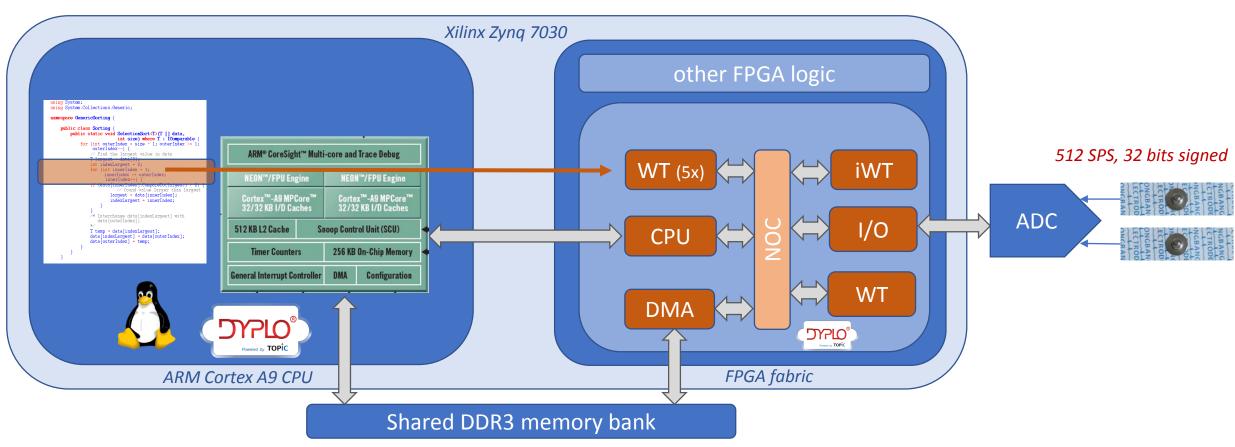






Programming model



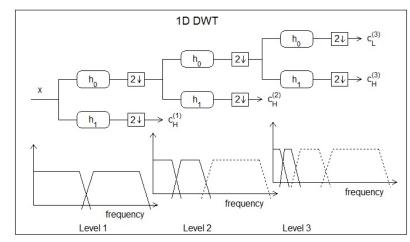


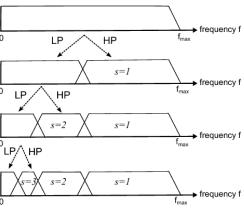
DESIGN AUTOMATION & EMBEDDED SYSTEMS



STEP 3: Understanding the problem

- ▲ 1-D Discrete Wavelet Transform
 - ▲ 4096 samples (double)
 - ▲ Apply high-pass and low-pass filter (8-point FIR)
 - ▲ Take only "even" results, 2048 samples each
 - ▲ The high-pass filter outcome is the result
 - A Repeat the process for the low-pass filter outcome
 - ▲ Results in 2 sets of 1024 samples
 - ▲ Repeat until only 1 sample remains
- ▲ How well are your mathematical skills?
 - ▲ How can we get around this?











STEP 4: Make it synthesizable



```
int pl = (int) tp.size();
while (L > pl) {
 tp.insert(tp.end(), t.begin(), t.end());
 pl = (int) tp.size();
t.insert(t.begin(), tp.begin() + pl - L + 1, tp.begin() +
pl);
t.insert(t.end(), tp.begin(), tp.begin() + L - 1);
DVec yl = conv(t, h); // lowpass filtering
DVec yh = conv(t, g); // highpass filtering
```

- •Use of std::vector<>
- -Use C arrays instead
- Moving data blocks around
- -Write data into output array on the correct location
- Inefficient code
- -Don't calculate what you don't need
- Too flexible
- -Fixed size input and output
- Constant filter coefficients

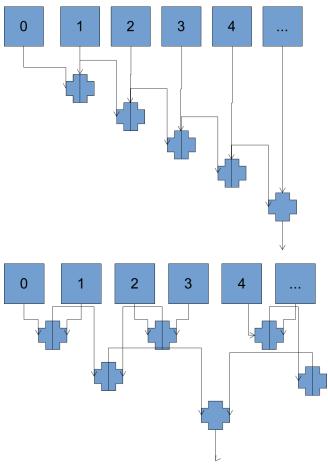
DESIGN AUTOMATION & EMBEDDED SYSTEMS





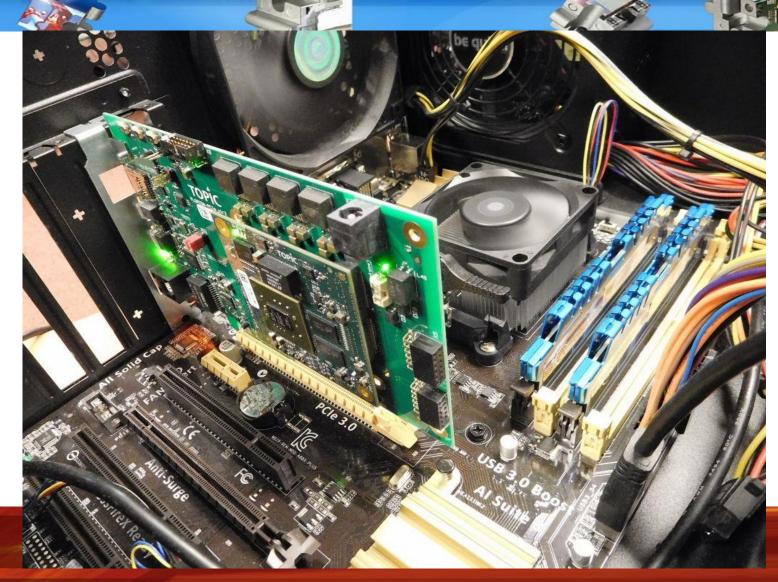
STEP 5: Speed it up

- Code describes a "linear" flow
- ▲ But a lot of operations can run in parallel
- ▲ Tools can do a lot, but need to "guess" your intentions
- ▲ Move code around (inline) to make intentions clear
- Reduce dependencies, usually by inserting "memory"





STEP 6: Test and Evaluate



DESIGN AUTOMATION & EMBEDDED SYSTEMS





Final results





	Original C++	Improved C++	FPGA
Number of cores	2		2
Time per step (µs)	8000	2000	363
Power consumption (mW)	400 mW		200 mW
Energy per step (mJ)	3200	800	72.6

▲ Notes:

- Placed two instances in the FPGA, hence "2" cores
- Power consumption measured at the 1V "core" power supply line
- The (improved) CPU implementation uses 11x more power







Interesting facts





- ▲~180 million double precision floating point MULT-ACC operations / second
 - ▲ Per wavelet data set 65536 MAC operations
 - ▲ 5 wavelet transforms concurrently
- ▲1 WT uses ~ 15% FPGA resources
- ▲ Acceleration from ~2ms to ~180us per WT
- ▲ Code manipulation gives best C-2-HDL optimizations, not just the directives
- ▲ Maintaining double precision floating point on FPGA (!!!)

- ▲ Data transfer bandwidth from CPU to FPGA becomes a dominating factor
- Remarkably fast implementation cycle by a non-FPGA embedded programmer
 - ▲ 4 days to make C++ code suitable for synthesis
 - ▲ 1 day of optimizing for synthesis to reach performance goal
 - ▲ Have a look what a WT IP license costs ...
- Starting point:standard mathematical C-models
 - ▲ Straight from the Internet, including test benches













Materiaalweg 4, 5681 RJ BEST, the Netherlands



+31 499 336979



www.topic.nl



dirk.van.den.heuvel@topic.nl



