nano-tera.ch
swiss scientific initiative in health / security / environment systems

IcySoC RTD 2013

FNSNF

Precise and Approximate Logarithmic Number Units shared in a Multi-Core Cluster

PULP Parallel Ultra Low Power

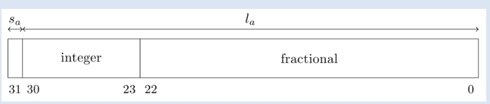
Michael Gautschi, Michael Schaffner, Frank Gurkaynak, Luca Benini

Integrated Systems Laboratory, ETH Zurich, Switzerland

1. The logarithmic number system:

The logarithmic number system (LNS) can be used to exploit a larger dynamic range. The format has its pros and cons, but leads to attractive results. We have developed a Logarithmic Number Unit (LNU) which achieves IEEE single precision and can be used as an alternative to a FPU.

LNS Number Representation:



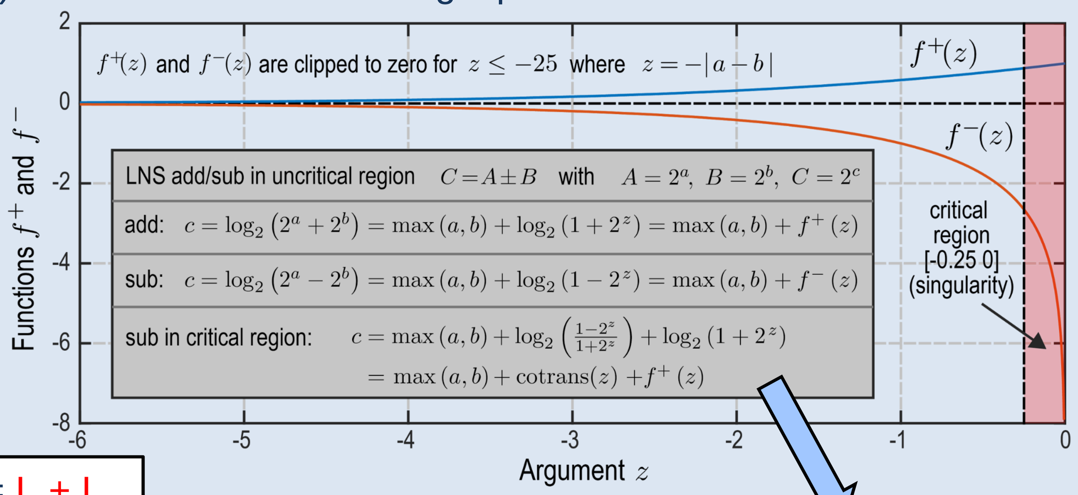
$a = (-1)^{Sa} * 2^{la}$

Simple operations in LNS:

$l_{mul} = \log_2(x*y) = \log_2(2^{lx} * 2^{ly}) = l_x + l_y$

$l_{div} = \log_2(x/y) = \log_2(2^{lx} / 2^{ly}) = l_x - l_y$

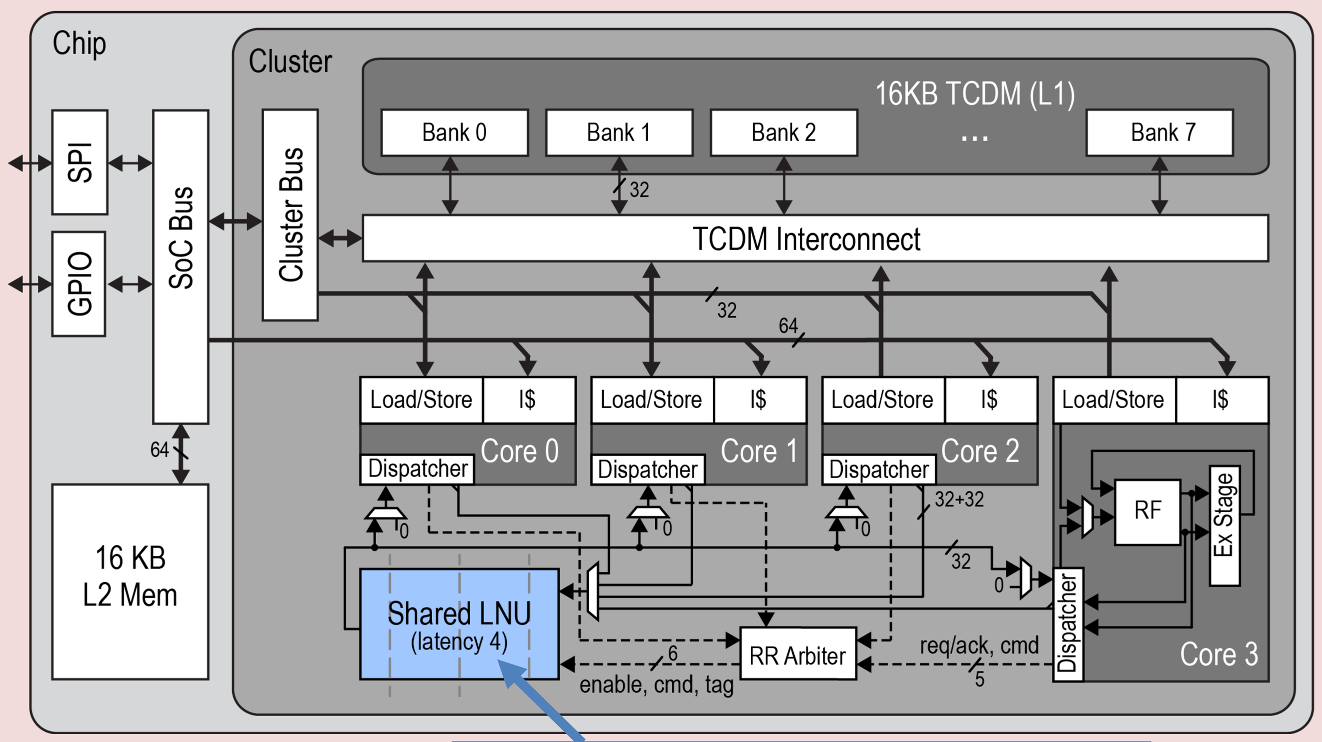
$l_{sqrt} = \log_2(x^{1/2}) = \log2(2^{lx/2}) = l_x > 1$



ALU of a core → Logarithmic Number Unit

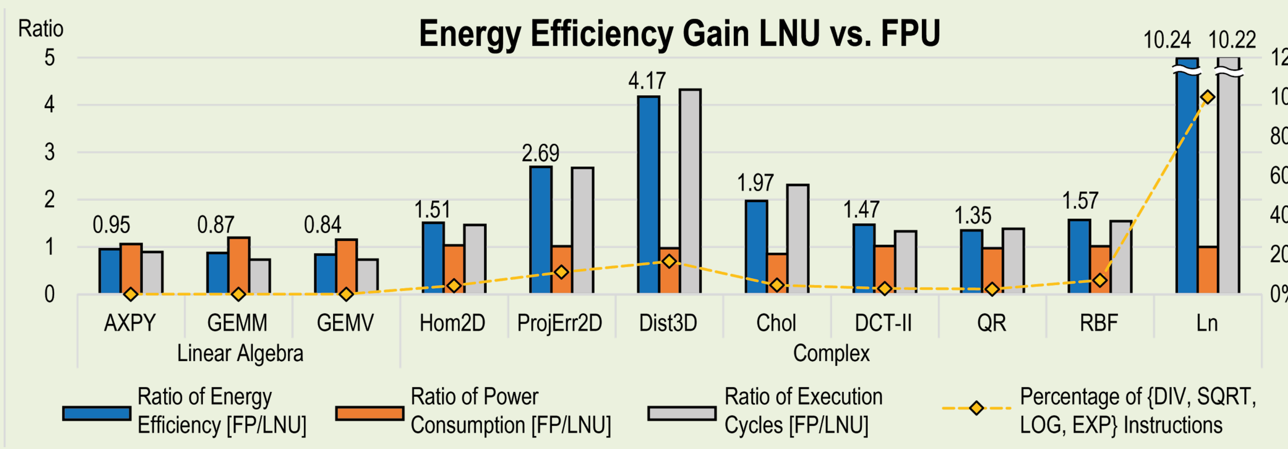
2. Sharing a LNU in a cluster of simple cores:

- 1 LNU shared among 4 cores
- Round Robin arbiter for fair sharing
- FP- MUL/DIV can be computed in the integer units of the cores
- SQRT support for free (shift)

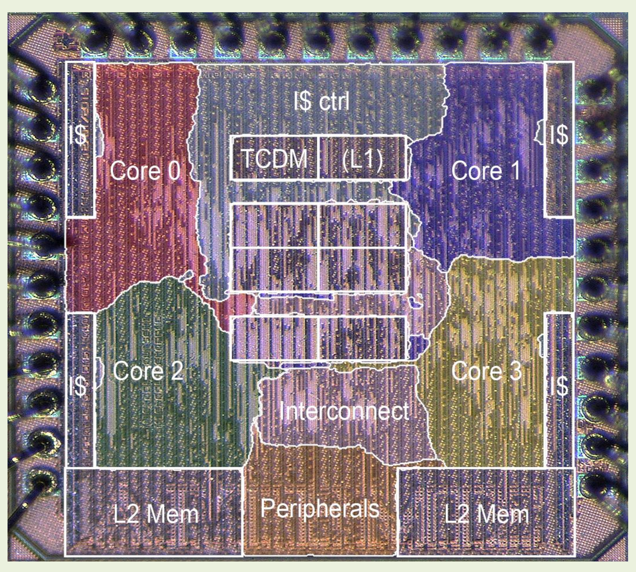


Shared Logarithmic Number Unit

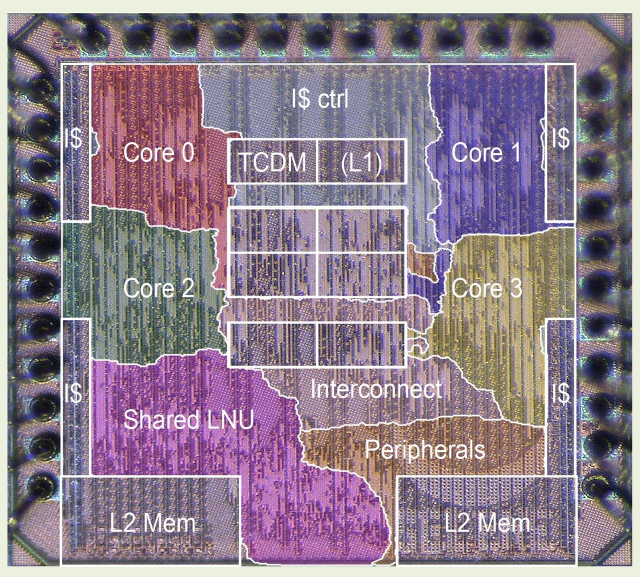
3. LNU vs FPU Comparison: [3]



Private FPU:



Shared LNU:

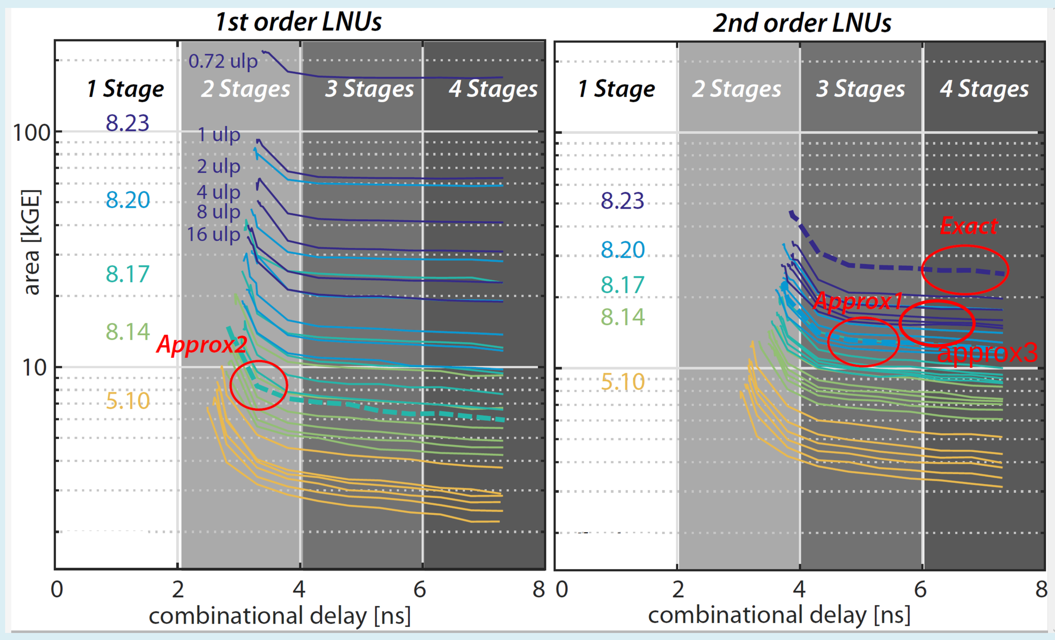


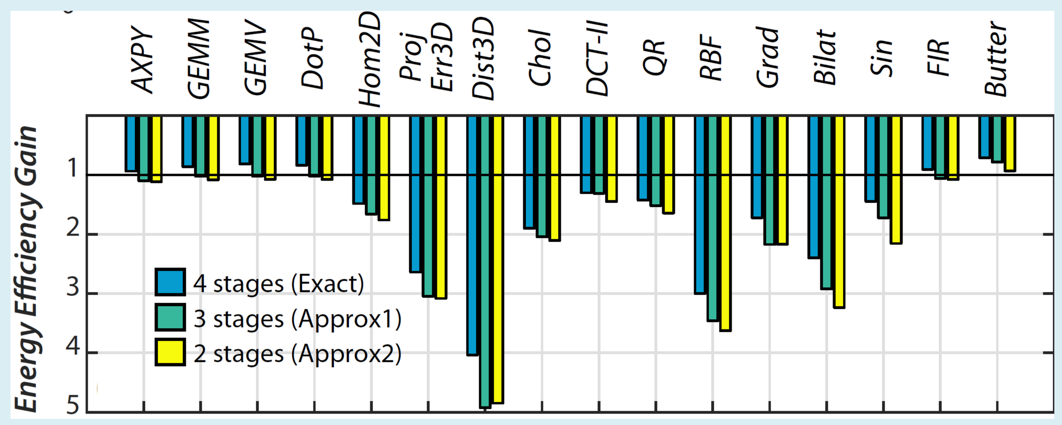
- LNU up to 4x more energy efficient than FPU when computing complex kernels.
- 1 LNU can be efficiently shared in a cluster of four processing cores.

Implementation Details		Private FPU[3]	Shared LNU[3]	ELM [1,2]
Technology		65nm LVT	65nm LVT	180nm
max speed [MHz]		374	337	125
max. Throughput [GFLOPS]		1.1	0.9	0.084
Power @100MHz, 1.2V, 25°C [mW]		41.84	44.0	-
Leakage @ 1.2V, 25°C [mW]		2.823	3.019	-
Precision (max err) [ulp]		0.5	0.478[1]	0.454[1]
avg. lnu/fpu utilization		0.21	0.37	-
Total area [kGE]		719	749	-
Single core area [kGE]		51.1[2]	44.5	-
Instruction support		**Private FPU**	**Shared LNU**	**ELM [1,2]**
Latency add/sub/casts	hw	2/2/2	4/4/4	3/3(4)/ -
Latency mul/div/sqrt [3]	hw	2/-/-	1/1/1	1/1/1
	sw	-/62/56	-/-/-	-/-/-
Latency exp/log [3]	hw	-/-	4/4	- / -
	sw	51/85	-/-	-/-

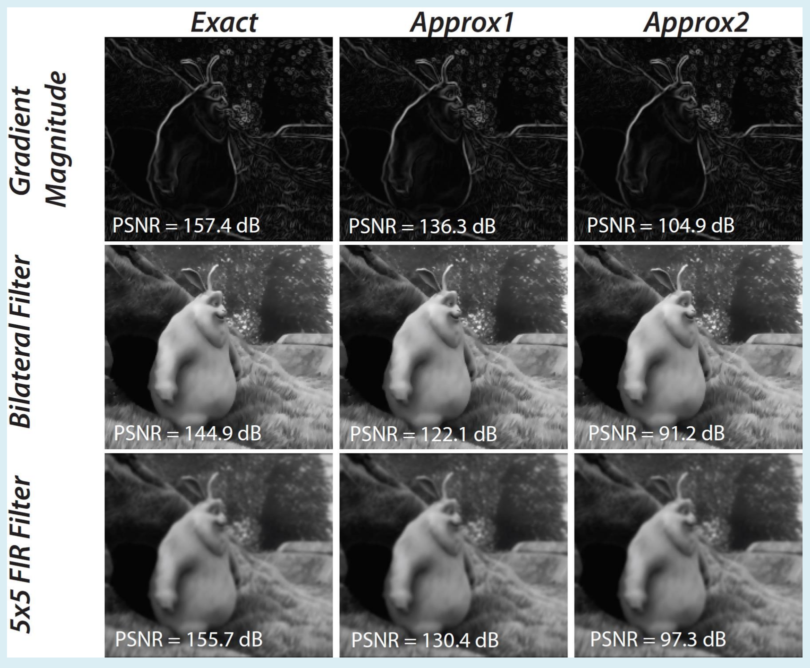
4. Evaluation of approximate LNUs: [4]

- Small errors and approximation can be tolerated in a lot of applications, for example in image processing
- The area of the LNU increases with the precision requirements.
- Approximation can be done by:
  - Reducing the bit width of the interpolators
  - Pruning lookup tables
- Tolerating errors leads to smaller LNUs, and smaller delay
  - Further allows to decrease the number of pipeline stages!





Approx2 is 24% more energy efficient than the exact LNU due to:
- 2 cycles latency instead of 4
- 28% faster execution
- Only 6% higher power consumption

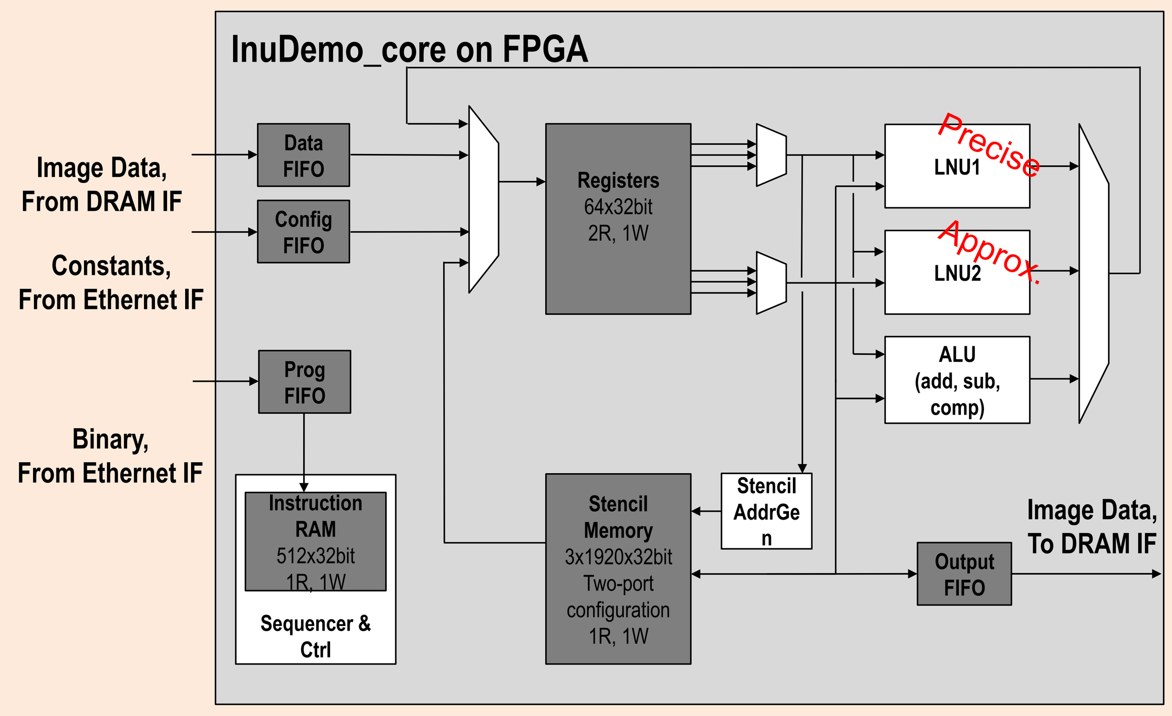


Applications:
- Gradient Magnitude:
  - Sobel Filter
  - Edge detection
- Bilateral Filter:
  - Nonlinear, edge preserving
  - noise.-reducing smoothing filter
- 5x5 FIR Filter:
  - Smoothing
  - Blurring

=> No visible, discernible errors

5. LNU Demonstrator:

- Image processing using approximate computing
- Demo Core with two LNUs
  - 64x32b registers
  - ALU to handle LNS MUL/DIV/SQRT
  - Precise or approx. LNU can be selected to compute ADD, SUB, EXP, LOG, Casts
  - Stencil memory for input image
  - Output written to frame buffer

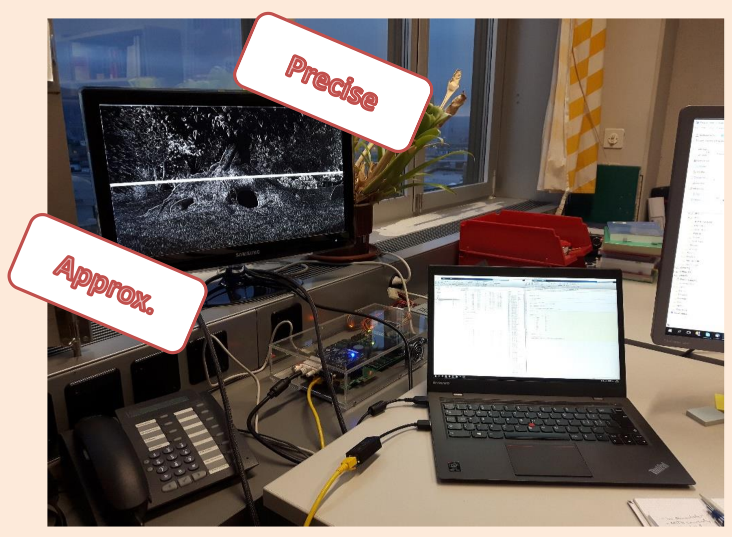


- Implemented on FPGA
  - Altera Stratix IV
  - 40 MHz
  - 113 kGE

LNU 1: (Single precision)
- 8.23 0.72 ulp
- 31.1 kGE

LNU 2: (Approx3)
- 8.23 16 ulp
- 21 kGE

Gradient Magnitude:



Mandelbrot:

6. References:

[1] "The European Logarithmic Microprocessor", J.N Coleman et.al, 2008
[2] "ROM-less LNS, R.Che Ismail and J.N Coleman", 2011
[3] "A 65nm CMOS 6.4-to-29.2 pJ/FLOP@ 0.8V shared logarithmic floating point unit for acceleration of nonlinear function kernels in a tightly coupled processor cluster", M. Gautschi et. al, ISSCC 2016
[4] "Accuracy and Performance Trade-offs of Logarithmic Number Units in Multi-Core Clusters", M. Schaffner et. al, ARITH 2016