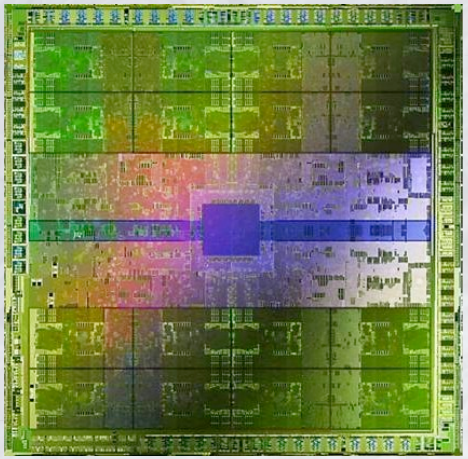
OPEN ISSUES IN HETEROGENEOUS SYSTEMS

LaME-in-Chief: David F. Bacon

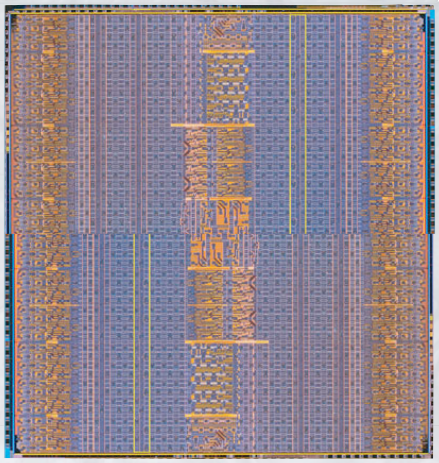
IBM Research

The Heterogeneous Era

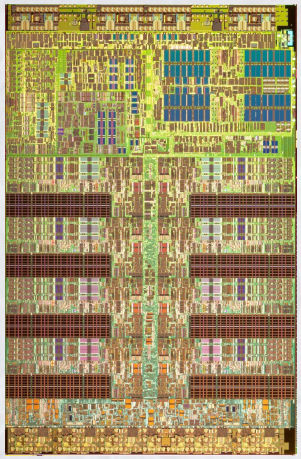


GPU

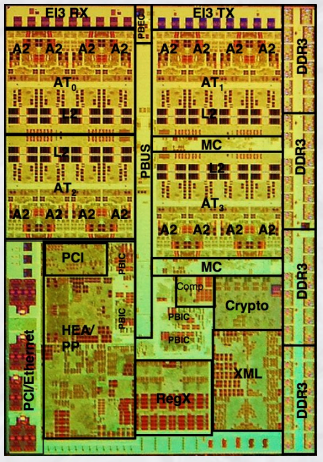
Tilera 64



FPGA

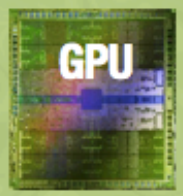


Cell BE



IBM PowerEN

What is Performance?

	Chip	flops / cycle	freq (GHz)	Gflops (peak)	Gflops / watt	Mflops / $	Mflops / watt/$
	Intel Core i7	32	3.2	102	0.8	70	0.7
	AMD 9270	1600	0.75	1200	5.5	800	3.6
	Xilinx V5 LX330	1040	0.55	550	13.7	138	8.1

Peak Performance

Source: Brodtkorb et al, The State-of-the-Art in Heterogeneous Computing, 2010

PERFORMANCE, TAKE 2

	Chip	Gsamples / sec	Msamples / joule
	Intel Core 2	1.4	5
	nVidia GTX280	14	115
	Xilinx Virtex 5	44	1461

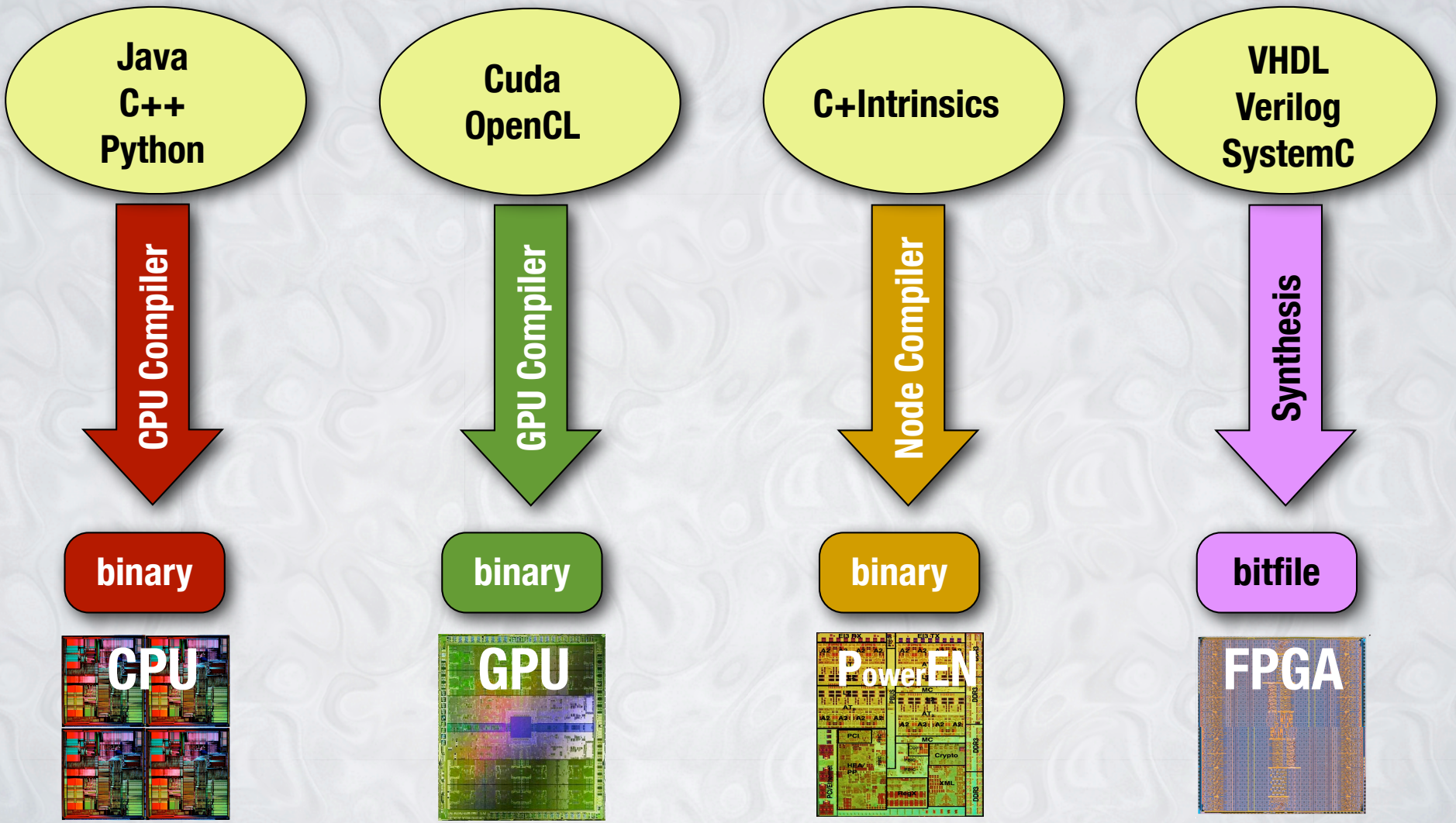
Actual Performance
(Random Number Generation)

Source: Thomas et al, A comparison of CPUs, GPUs, FPGAs and masssively parallel processor arrays for random number generation, 2009

Opportunities in Heterogeneous Systems

- Performance
  - 10-1000x speedups
- Efficiency
  - 10-100x improvement in ops/watt

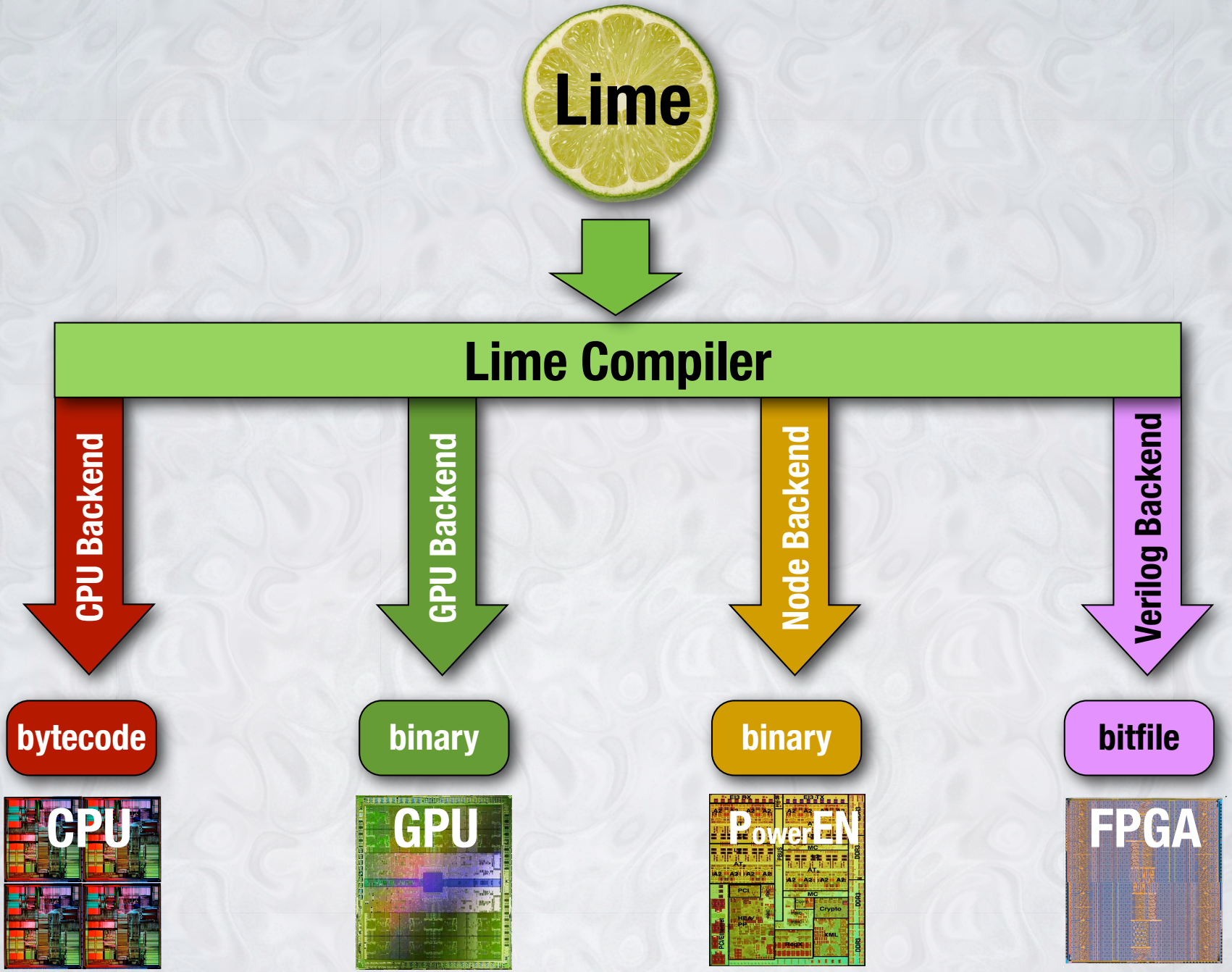
HETEROGENEOUS PROGRAMMING TODAY



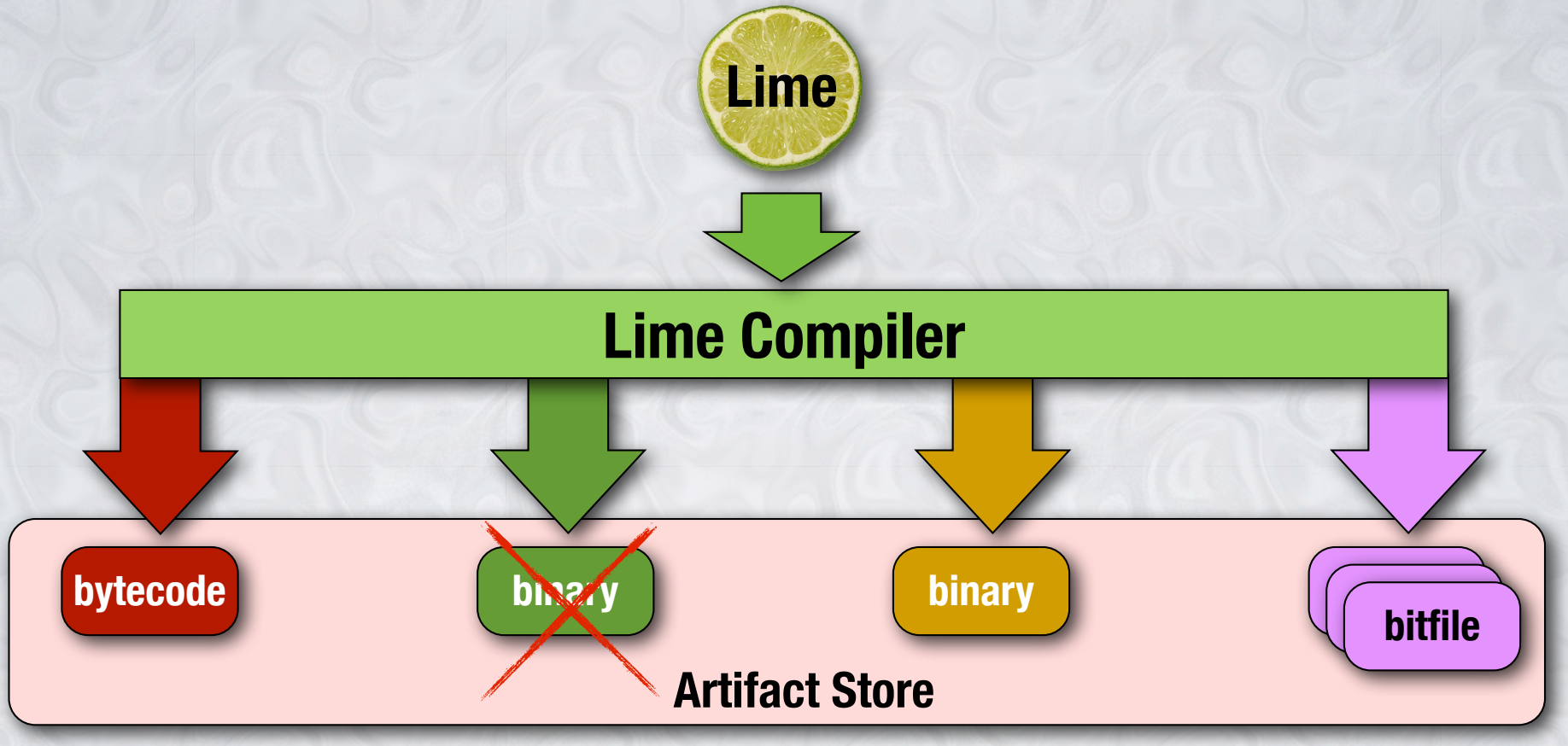
Open Issues in Heterogeneous Systems

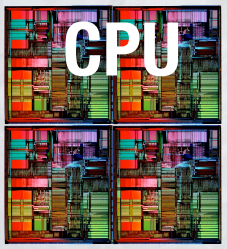
1. Heterogeneous languages

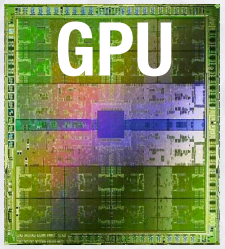
The Liquid Metal Programming Language

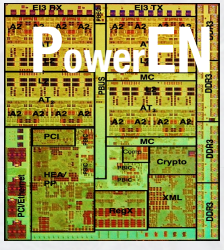


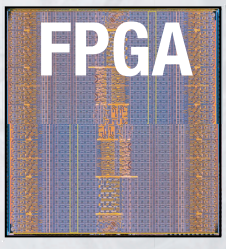
The Artifact Store & Exclusion







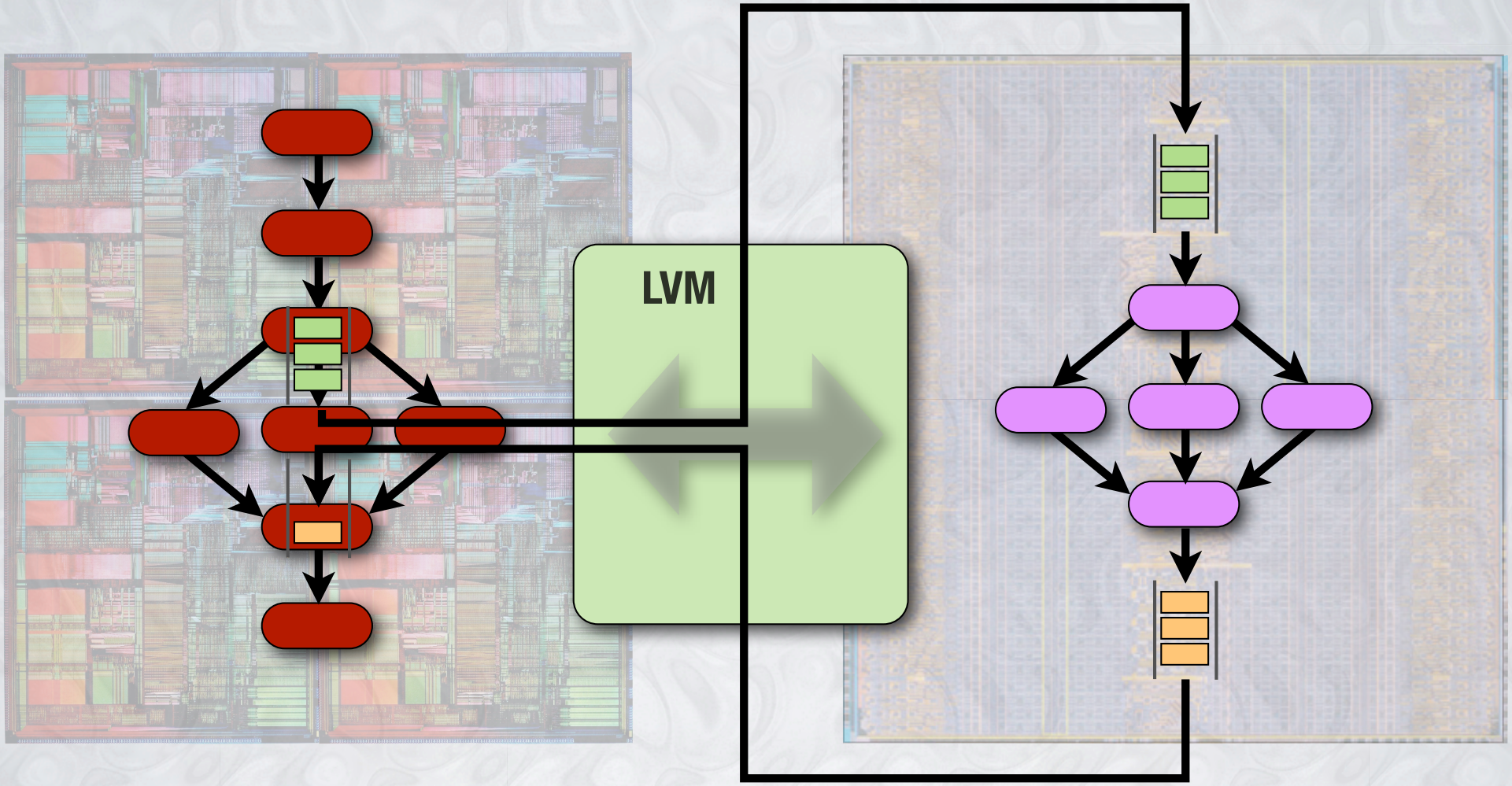




Open Issues in Heterogeneous Systems

1. Heterogeneous languages
2. Some programs can’t or shouldn’t be expressed

Execution, Communication, and Replacement



Open Issues in Heterogeneous Systems

1. Heterogeneous languages
2. Some programs can't or shouldn't be expressed
3. Data transfer is under flux and highly variable
    a. Ongoing debate and experimentation with coherent attach

Lies, Damned Lies, and FPGAs

- "It works"
  - A logic analyzer shows it meets the interface spec
- "They can be reprogrammed in a few milliseconds"
  - A few milliseconds + a few minutes to reboot
- "They have rich, high-speed I/O connections"
  - Which are very hard to talk to
- "...so they can interface to anything"
  - Which means they work with almost nothing
- "Synthesis times can be reduced with partial reconfig"
  - Partial reconfig is always available in Now + 7 months

Open Issues in Heterogeneous Systems

1. Heterogeneous languages
2. Some programs can't or shouldn't be expressed
3. Data transfer is under flux and highly variable
    a. Ongoing debate and experimentation with coherent attach
4. FPGAs have high bring-up costs (40-60% of a project)

Performance and User Experience

	Chip	Gsamples / sec	Msamples / joule	
	Intel Core 2	1.4	5	2 ½ minutes
	nVidia GTX280	14	115	16 seconds
	Xilinx Virtex 5	44	1461	5 seconds

Open Issues in Heterogeneous Systems

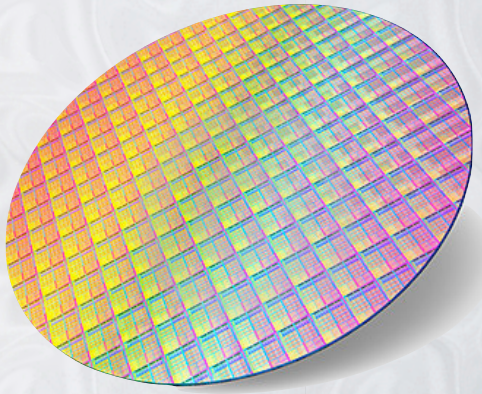
1. Heterogeneous languages
2. Some programs can't or shouldn't be expressed
3. Data transfer is under flux and highly variable
    a. Ongoing debate and experimentation with coherent attach
4. FPGAs have high bring-up costs (40-60% of a project)
5. Even if we hide complexity, we can't hide performance

Sharing Devices?

Open Issues in Heterogeneous Systems

1. Heterogeneous languages
2. Some programs can't or shouldn't be expressed
3. Data transfer is under flux and highly variable
    a. Ongoing debate and experimentation with coherent attach
4. FPGAs have high bring-up costs (40-60% of a project)
5. Even if we hide complexity, we can't hide performance
6. No virtualization (is this a bad thing?)

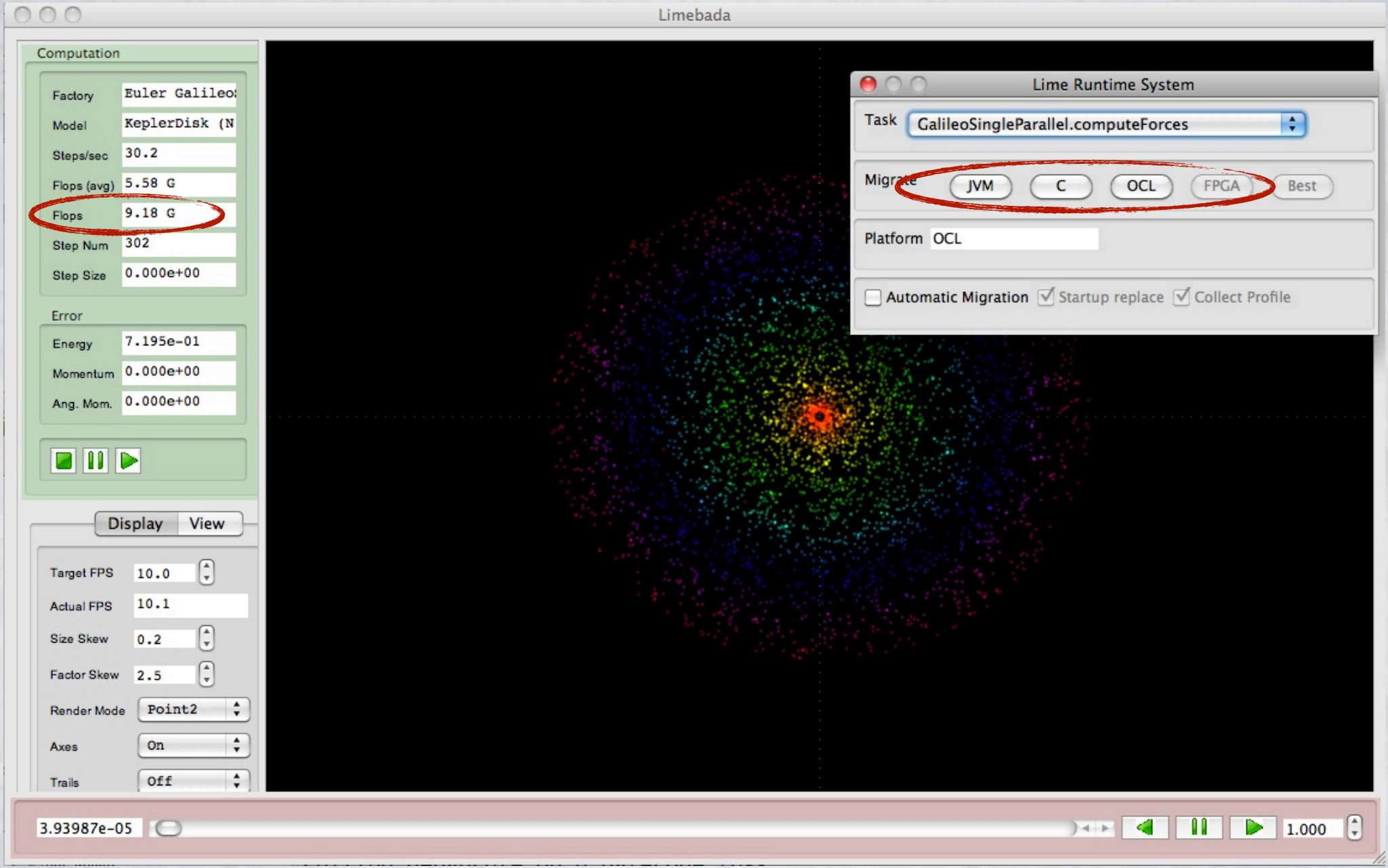
What's Our Metric?



Open Issues in Heterogeneous Systems

1. Heterogeneous languages
2. Some programs can't or shouldn't be expressed
3. Data transfer is under flux and highly variable
    a. Ongoing debate and experimentation with coherent attach
4. FPGAs have high bring-up costs (40-60% of a project)
5. Even if we hide complexity, we can't hide performance
6. No virtualization (is this a bad thing?)
7. No methodology for evaluation

But it Sure is Nice When it Works



DISCUSSION