A Cycle-accurate, Cycle-reproducible, multi-FPGA System for Accelerating Multi-core Processor Simulation

Sameh Asaad, Ralph Bellofatto, Bernard Brezzo, Chuck Haymes, Mohit Kapur, Benjamin Parker, Thomas Roewer, Proshanta Saha, Todd Takken, Jose Tierno

IBM T.J.Watson Research Center

Yorktown Heights, NY 10598

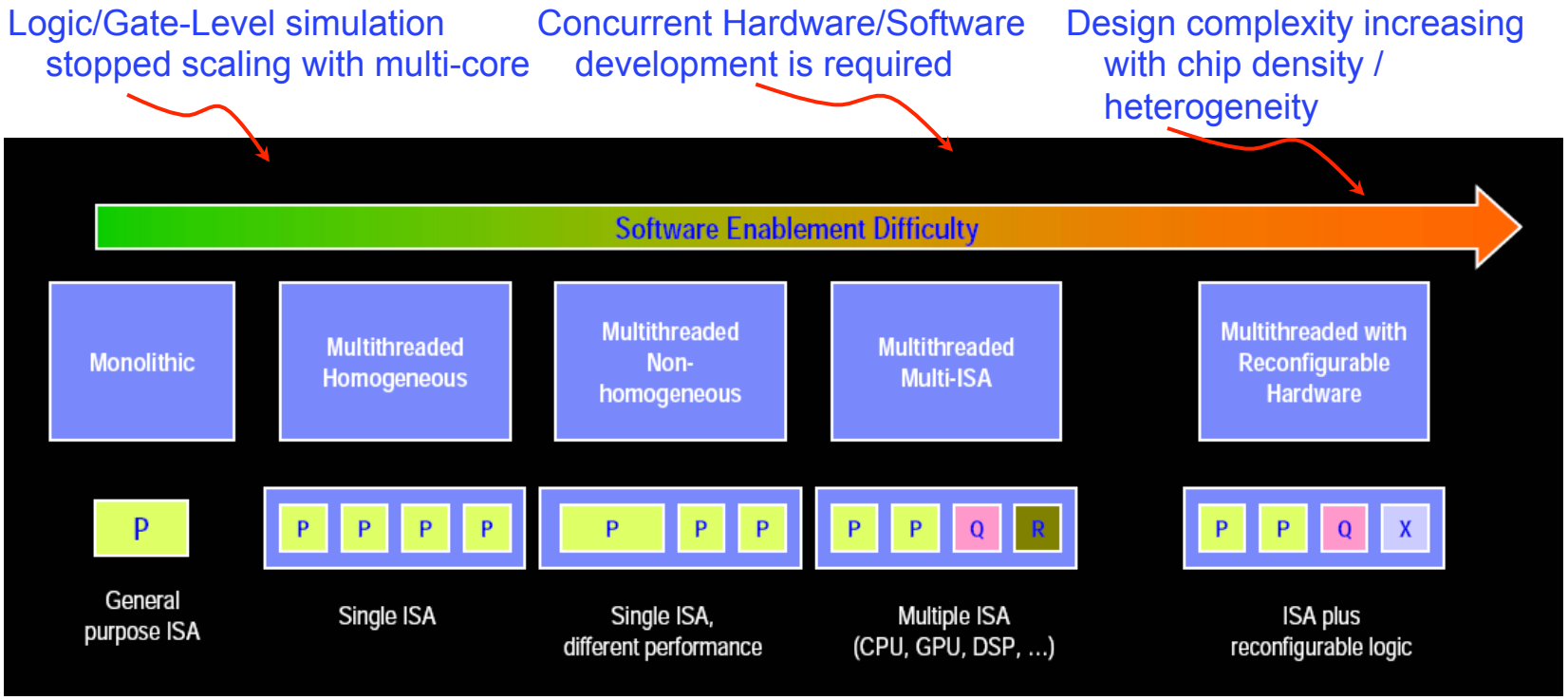
Feb 2012

Acknowledgments

- The Blue Gene/Q project, and consequently the Twinstar FPGA platform, has been supported and partially funded by Argonne National Laboratory and the Lawrence Livermore National Laboratory on behalf of the United States Department of Energy, under Lawrence Livermore National Laboratory Subcontract No. B554331

- Some aspects of this work were done in close collaboration with Xilinx, in particular on state read-back using the internal capture (ICAP) macro for Virtex-5 devices.

Increasing complexity of chip & system design is driving a growing need for emulation…

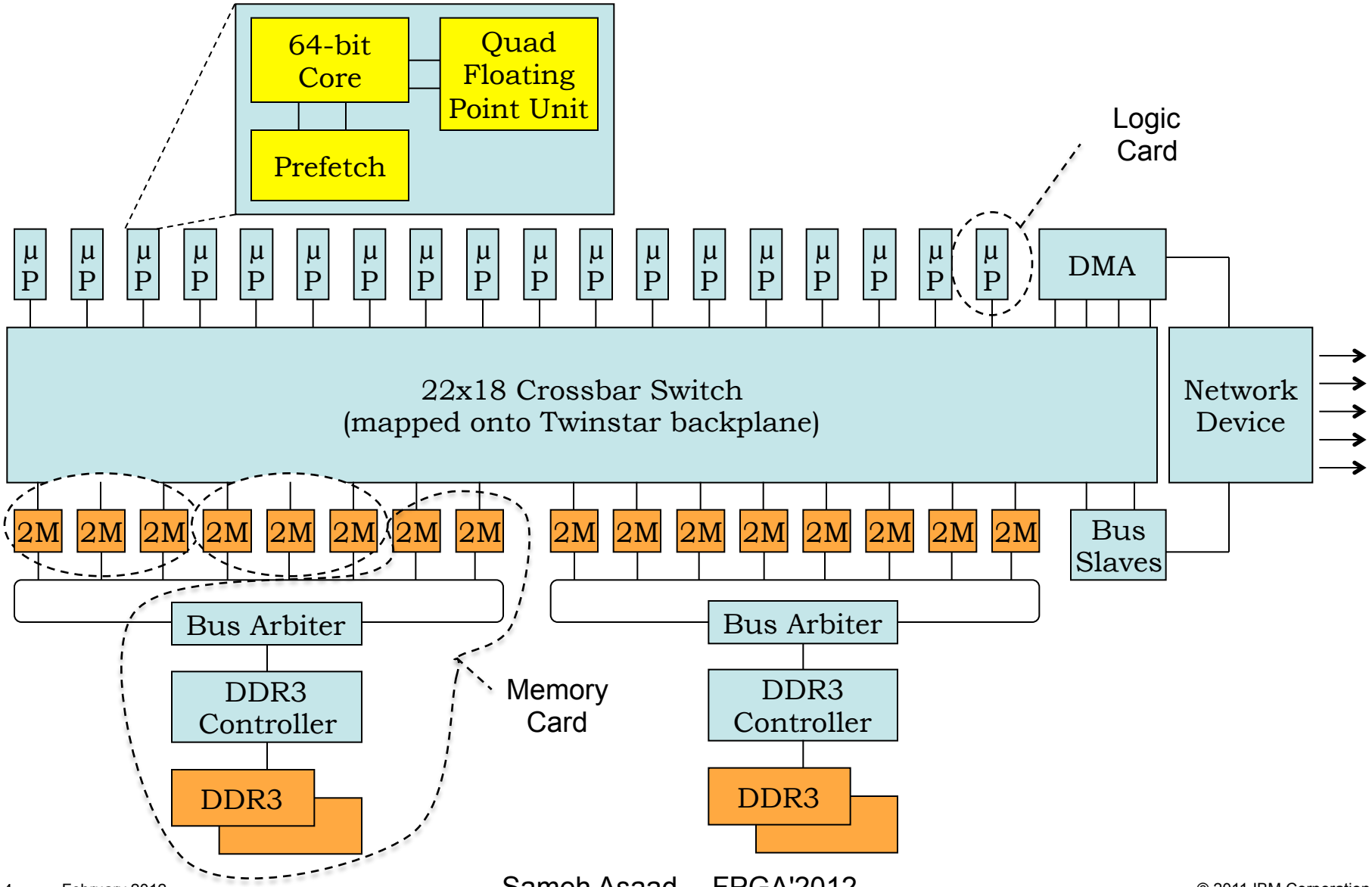


Mask costs increase with technology node →

Time-to-market pressure rising →

- *Paradigm of using current system to develop next generation is broken (verification, software porting/tuning..)*
- *FPGA emulation accelerates time-to-market and reduces development risk*

The Bluegene/Q processor architecture

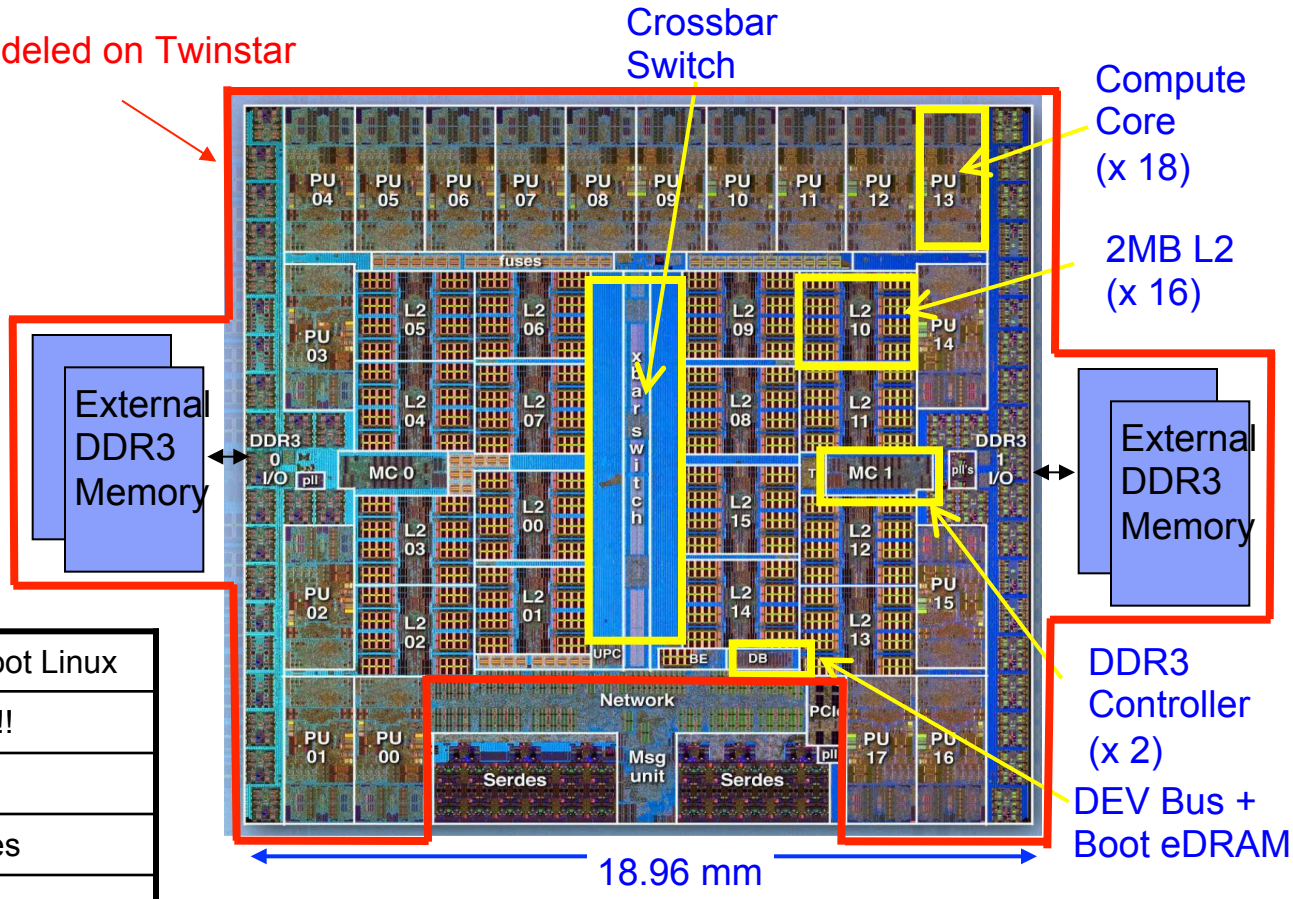


Modeling BG/Q compute node

- *Challenges include : large (18) number of cores, unique cross-bar architecture, and eDRAM*
- *All of the chip (except network unit) + ext. DDR3 memory modeled on FPGA (using 45 devices)*
- *FPGA platform runs at 4 MHz simulated processor clock speed (compared to 10 Hz in S/W)*

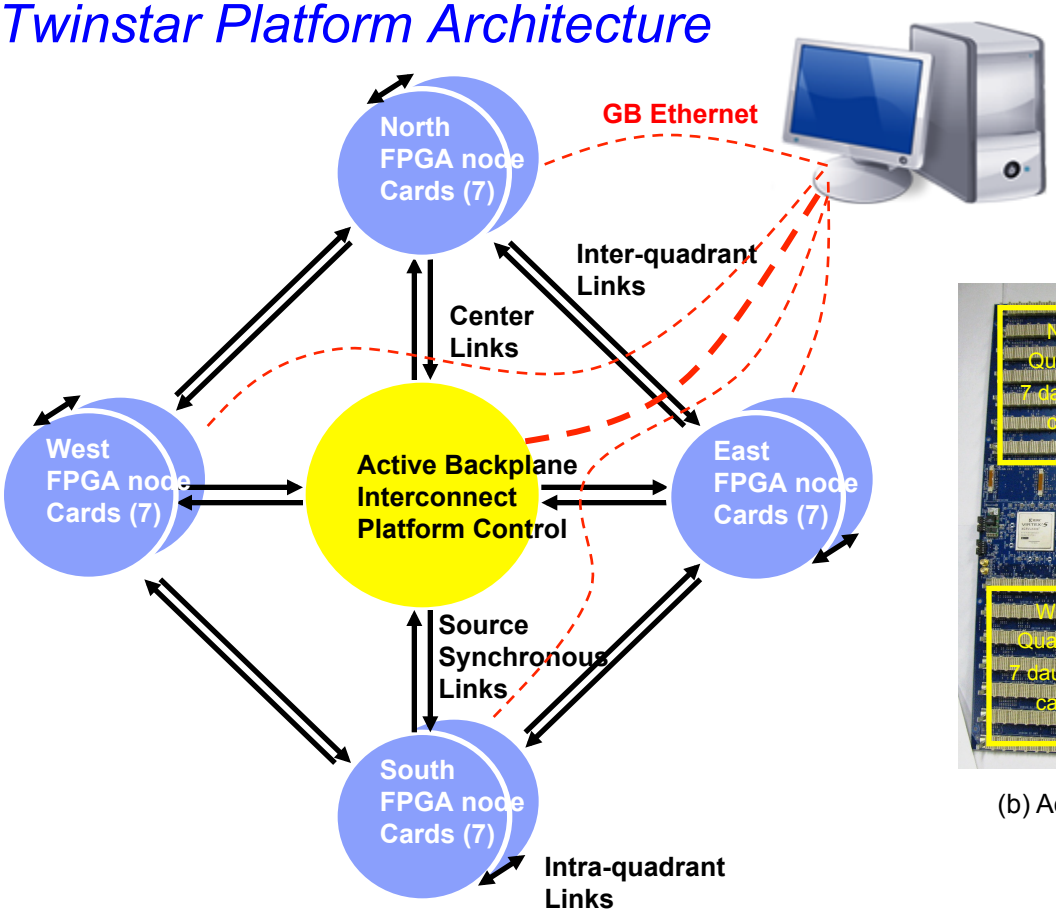
KEY CHIP FEATURES

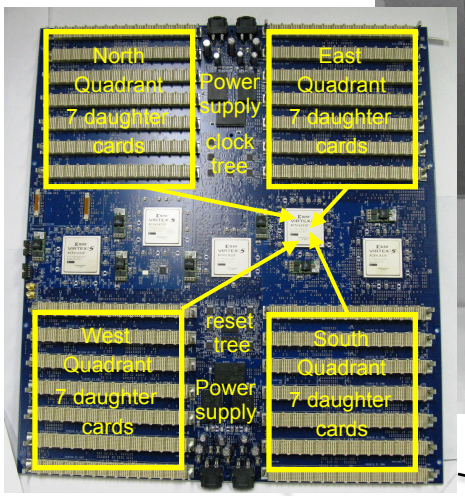
- 45-nm SOI technology
- 204.8 GFlops/sec @ 55W
- 17-way SMP
- 32 MB L2 cache (eDRAM)
- full crossbar switch
- Area: 360 mm^2



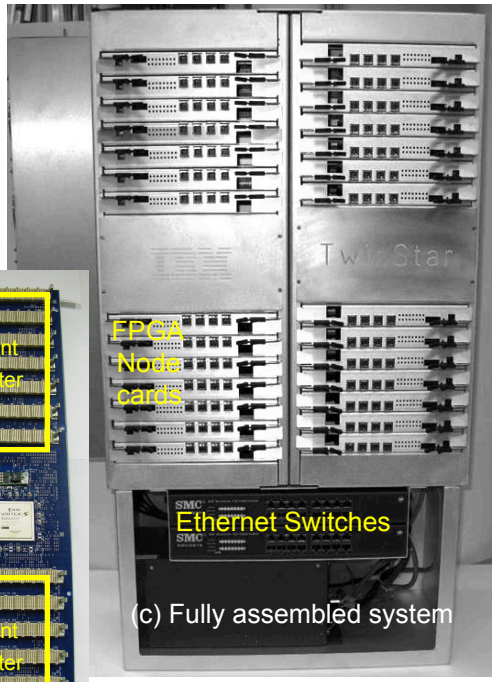
	Frequency	Time to boot Linux
Simulation	10 Hz	4.8 years !!
AWAN	1 KHz	17.6 days
Twinstar	4 MHz	6.3 minutes
ASIC	1600 MHz	1 second

Twinstar Platform Architecture

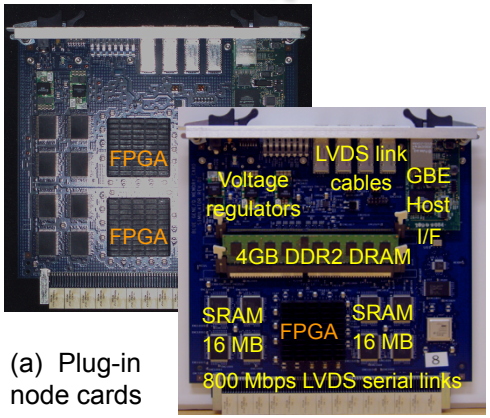




(b) Active Backplane



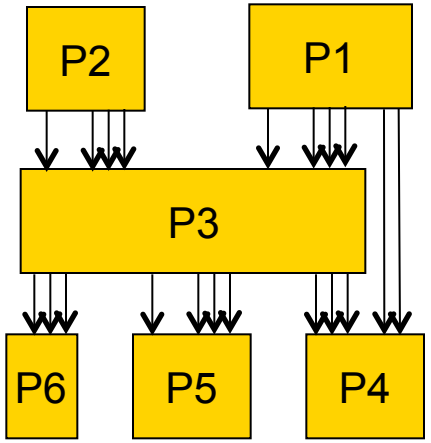
(c) Fully assembled system



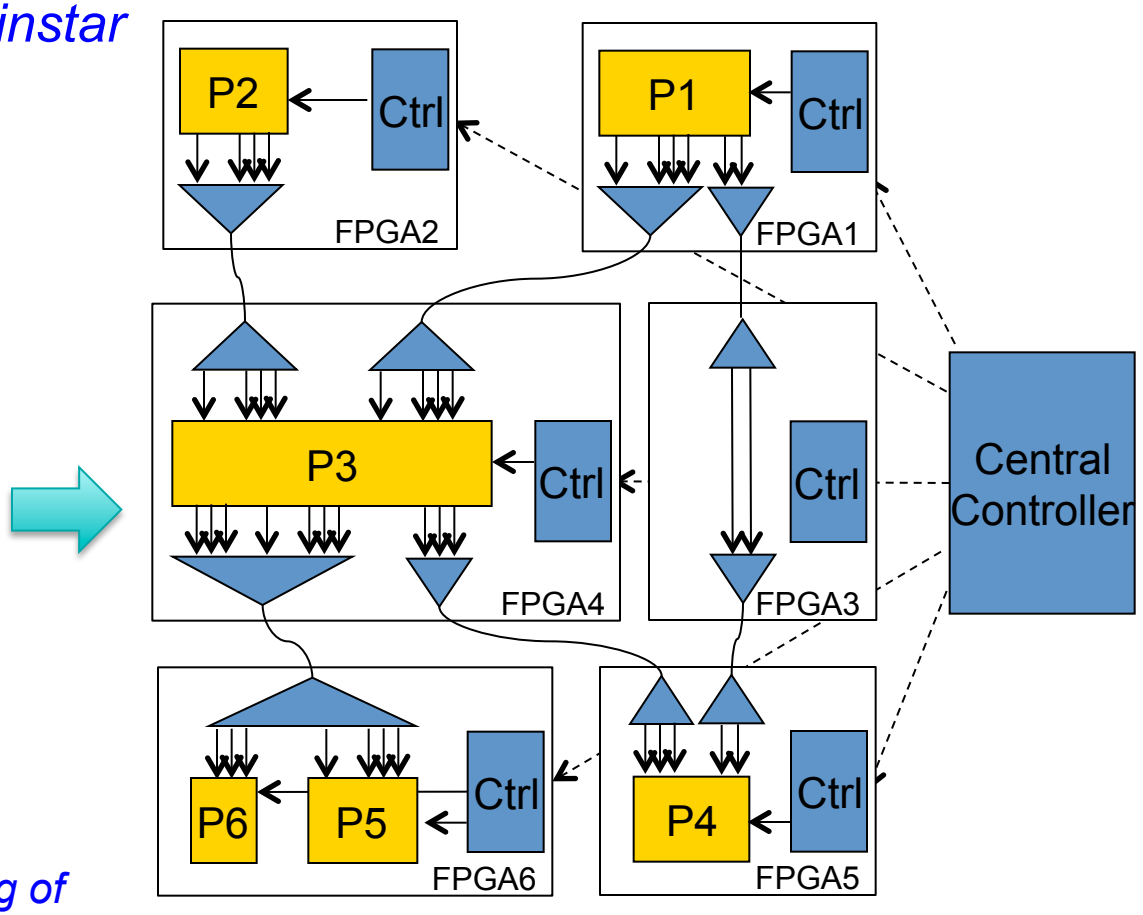
(a) Plug-in node cards

- *Tightly-coupled system of FPGAs*
- *All devices are synchronized for cycle-accurate, cycle-reproducible execution*
- *High-bandwidth GBE links to host for fast uploading of waveform data*

Partitioning & mapping onto Twinstar



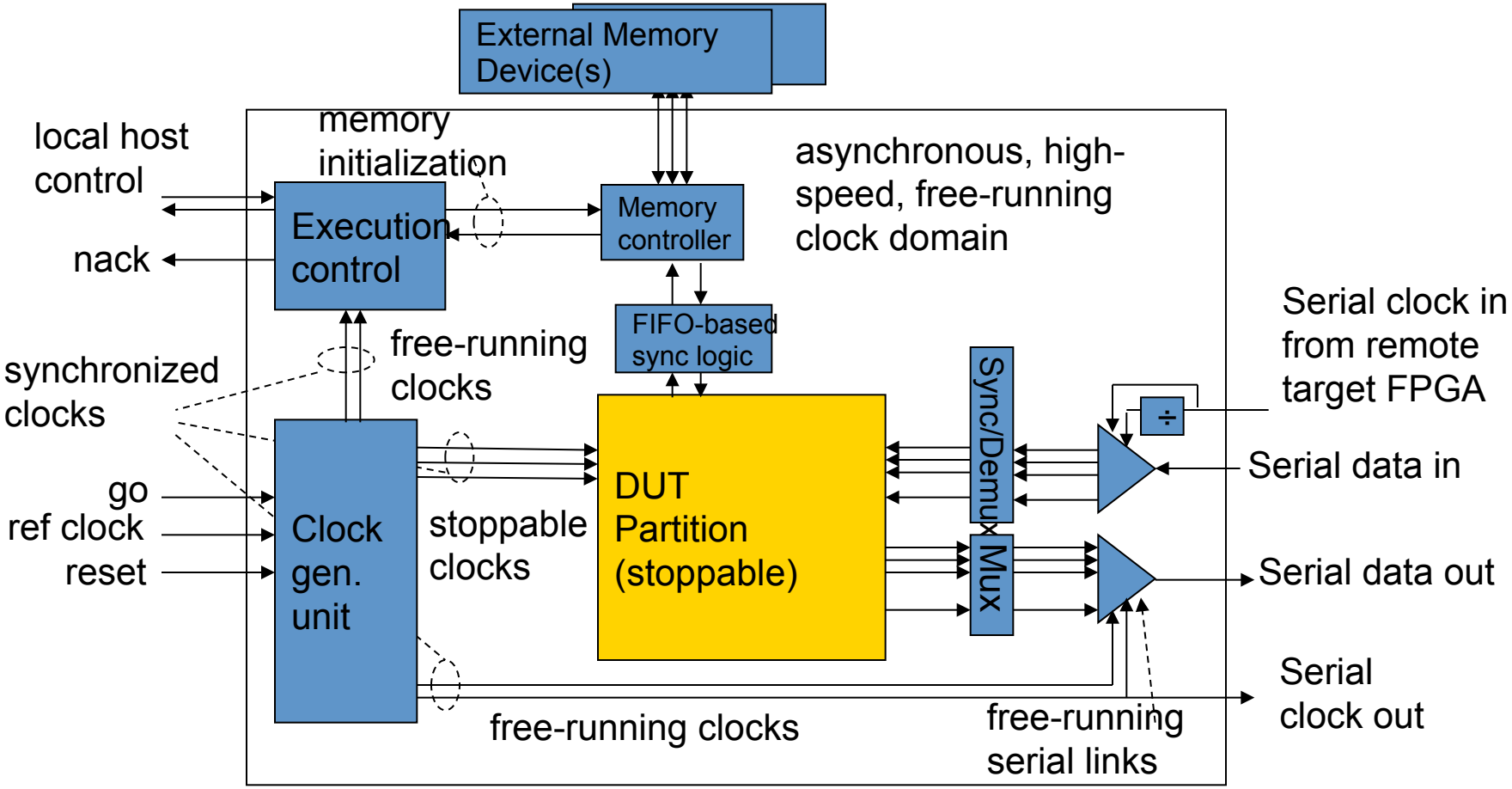
Logical (Chip) View



Physical (FPGA) View

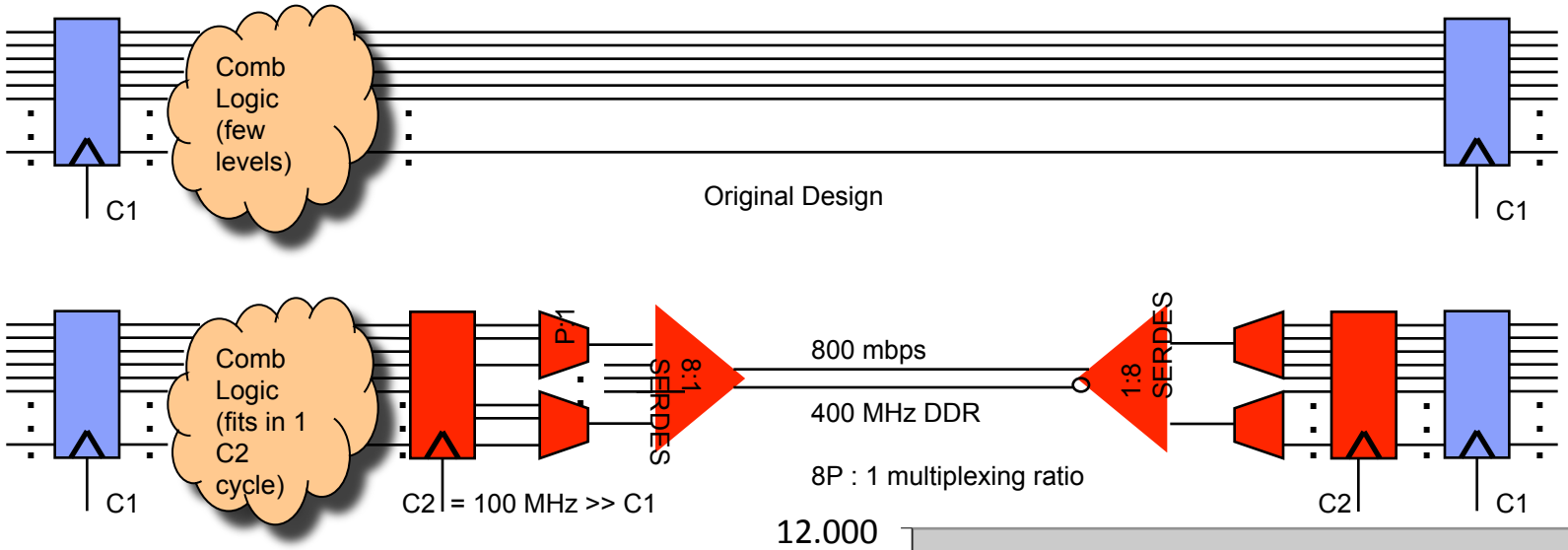
- *One-to-one and many-to-one mapping of chip partitions onto individual FPGA devices*
- *Serializer-deserializer communication links inserted between partitions to fit physical communication network*
- *Through-channels using one-hop if no direct connection exists between source & destination*
- *Special treatment for clock and reset networks*
- *Chip partitions are stoppable, infrastructure logic is free running*

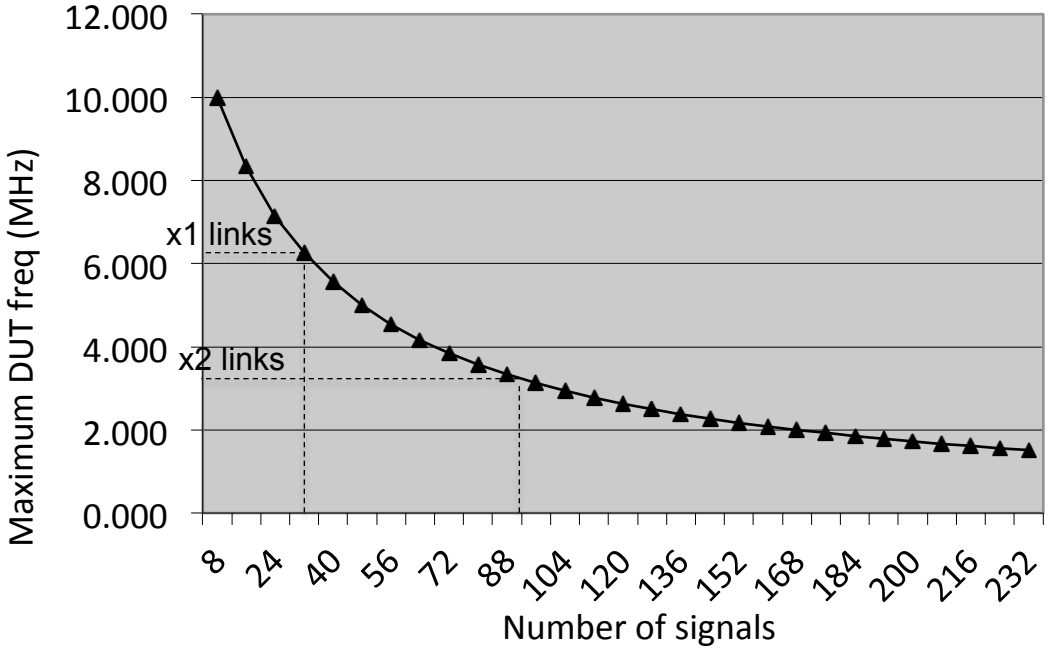
Cycle-reproducible execution using stoppable & free running clocks



- *DUT partition can be stopped at any cycle to examine / alter its state*
- *Infrastructure logic is free-running to maintain serial link training & memory refresh*
- *Clock generation circuit (replicated on each FPGA) ensures glitch-free stop & re-start*

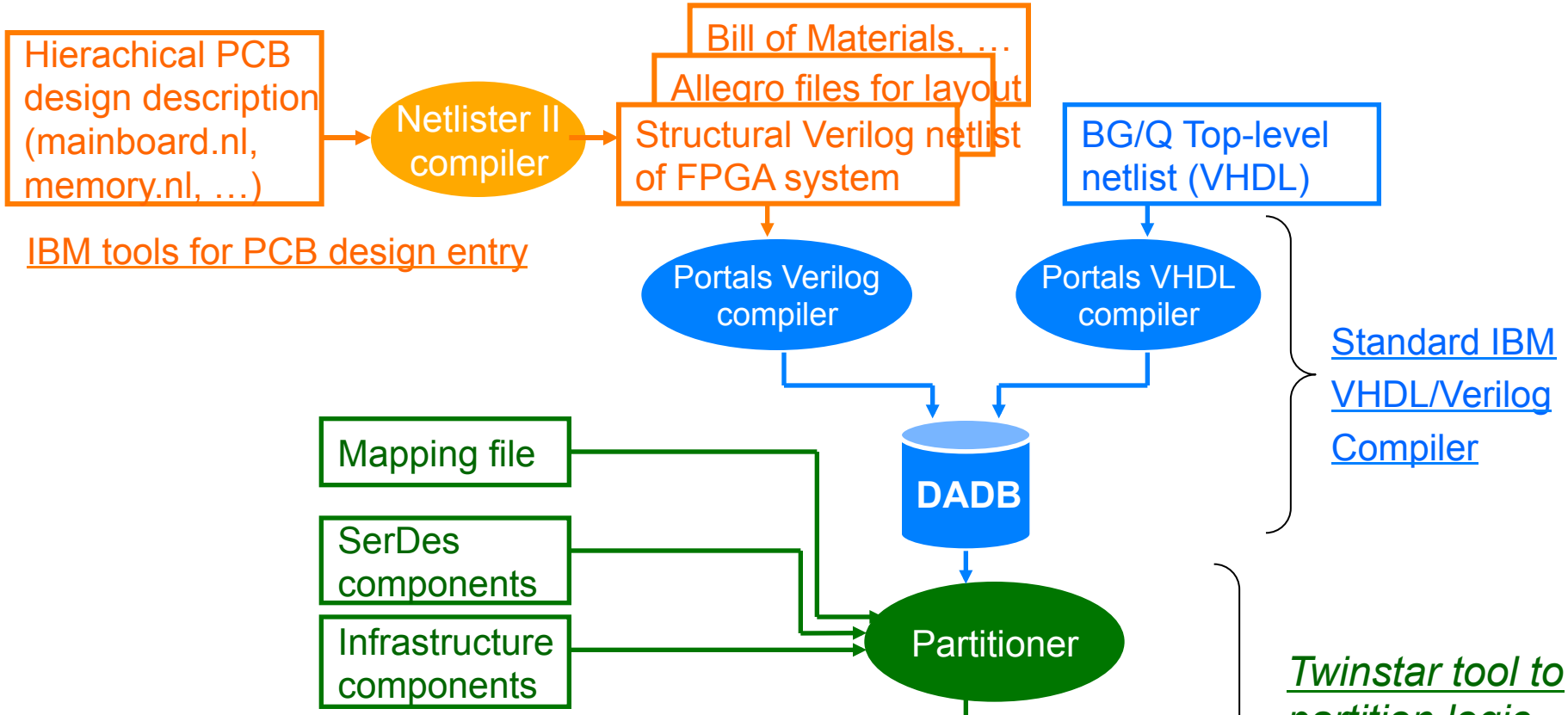
Serial Link design for multi-fpga partitioning





- *Transparent design from a DUT perspective (implements a wire)*
- *Represents the limiting factor in system emulation performance*

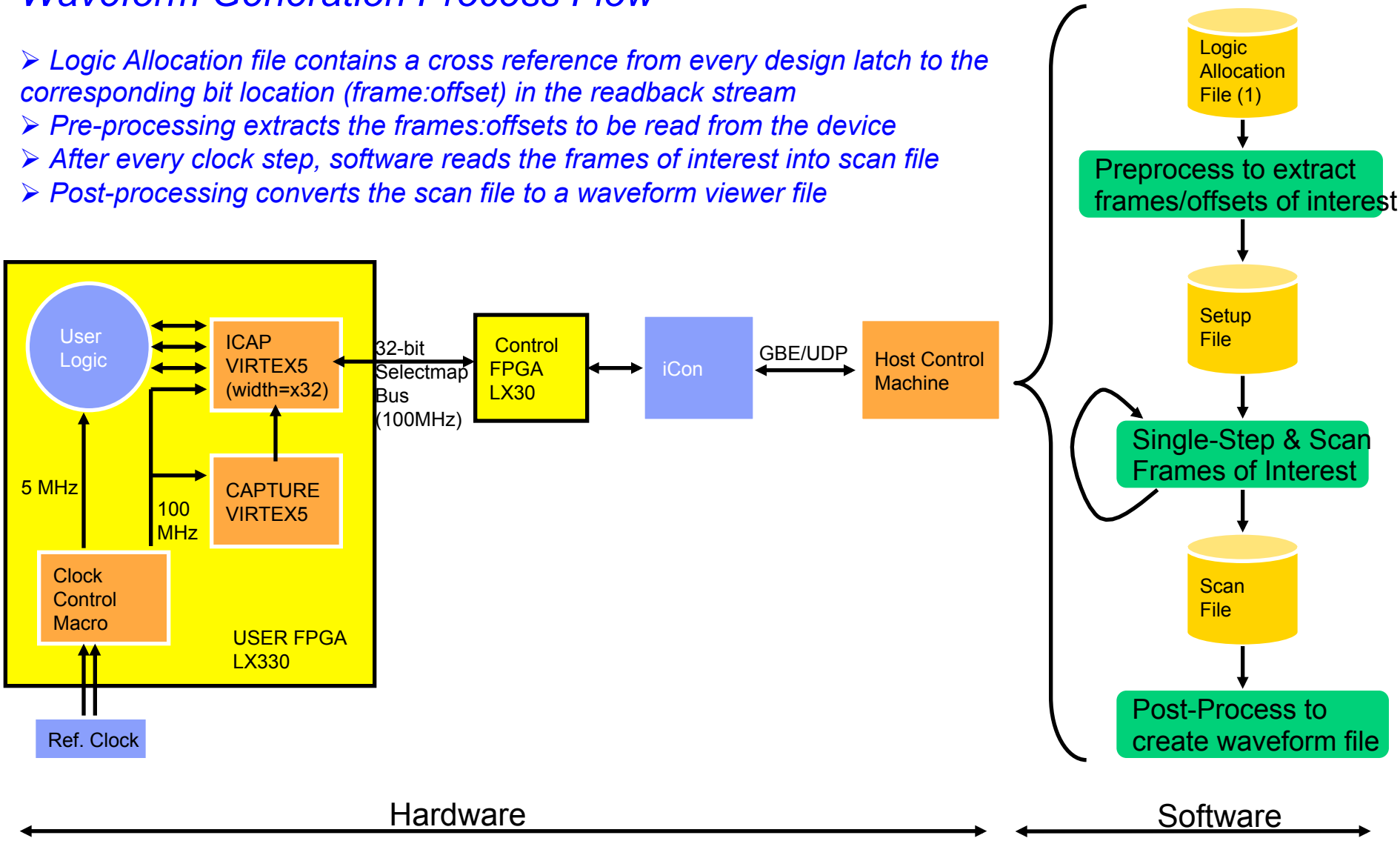
Design Methodology and Tools for Twinstar



- *Tool automatically generates top level netlist for each FPGA in the system*
- *Automatic mapping to external memory elements*
- *Each FPGA is synthesized, placed and routed in parallel*

Waveform Generation Process Flow

- *Logic Allocation file contains a cross reference from every design latch to the corresponding bit location (frame:offset) in the readback stream*
- *Pre-processing extracts the frames:offsets to be read from the device*
- *After every clock step, software reads the frames of interest into scan file*
- *Post-processing converts the scan file to a waveform viewer file*



Twinstar Impact on Bluegene/Q

Verification
Reduce RIT risk

1. **Over 30 chip-level bugs uncovered / fixed / validated**
   - Thousands of targeted verification tests ran (tens of billions of cycles)
   - New modes of operation in DD2 RIT only verified extensively using Twinstar : Transactional memory and thread-level speculation.

Performance Validation
Assure Quality

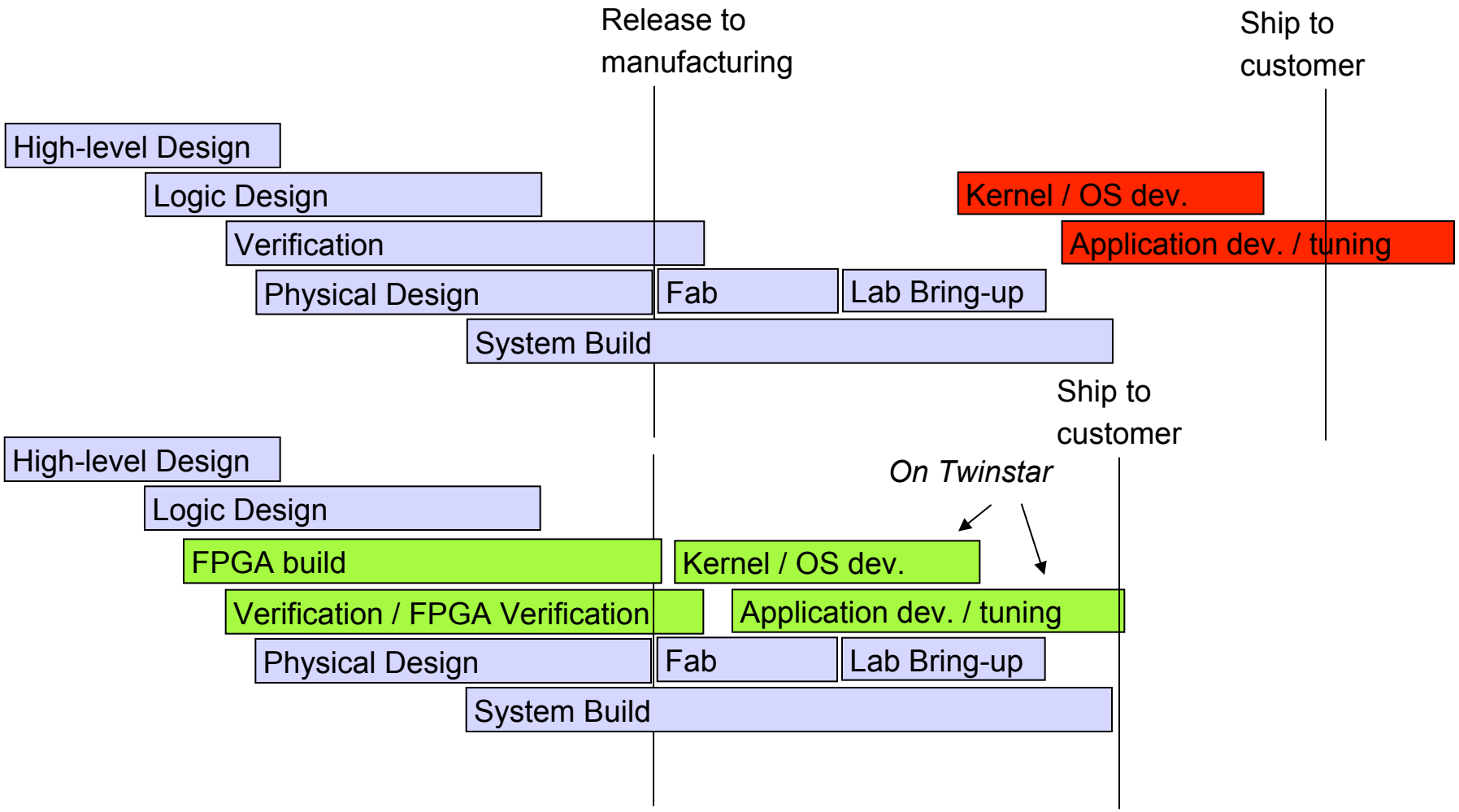
2. **Major Sequoia Benchmarks developed & tuned on Twinstar**
   - Major kernels of Sequoia Benchmarks have been executed to validate projected BG/Q node and system performance ahead of chip RIT

Software enablement
Time-to-market

3. **BG/Q kernel software developed using Twinstar**
   - Early development of kernel software for BG/Q before ASIC hardware bring-up
   - Development and tuning of key HPC codes for Bluegene/Q architecture

- *Key enablement of the above through cycle-accurate, cycle-reproducible accelerated simulation of BG/Q compute node*

Reducing time to market through FPGA platform



- *FPGA platform enables early kernel and application software development, reducing overall time to market*

Conclusion

- Twinstar provides 100,000x speed improvement over software simulation, while retaining cycle-accurate and cycle-reproducible behavior on a full-chip level simulation

- Platform enables new capabilities before hardware bring-up, not possible with conventional flow:
  - Vastly faster RTL verification at chip-level (e.g. memory coherence)
  - Early OS/ kernel software development
  - Application development / tuning for key HPC codes
  - Performance validation of Sequoia benchmarks

- Flexible building blocks enabling rapid construction of different instances, while retaining high-performance and low overhead.