Berkley vision vs. ERA proposal

Berkley vision	ERA proposal
Seven goals (and 11 bullet points) - Future computers must be effectively parallel; - It is possible to consider 1000+ cores; - Performance measurement for parallel computing should be re-evaluated and new metrics introduced.	- New rigorously defined computing paradigm based on GLM exploits automatically maximum possible parallelism of both: algorithm and hardware available and minimizes synchronisation complexity at the run time to reduce concurrency. This approach guarantees, together with flexibility of hardware configuration for the both: performance or reliability purposes. Adjust available ERA hardware for maximum possible efficiency in the whole broad class of applications usually covered by semantically different architectures such as VLIW, SIMD, MIMD.
- Auto-tuning of software and hardware; - Human centric computing in multi-core; - Application of wide range of data typing.	- Software will be written in traditional way, and debugged on a standard systems; parallelization of the algorithm will be done by backward compilation of a sequential program, using GLM again for representation of task potential parallelism and fine-tuning of hardware resources available on the wafer, before and during application run.
- Three levels of parallelism should be pursued: task level, word level and bit level; - Parallel programs must be presented independently to the number of processors available - Limitations on features that reduce parallelism - OS functionality should be based on libraries and virtual machines.	- Task level parallelism is prerogative of run time system and heavily depends on workload and resources available during application run: therefore, a dynamic scheduling to optimize parallelism of tasks is exception not the rule in ERA. ERA as any other multi-core systems is dealing with substantial amount of data and program processing, dynamic support of dynamic parallelism should be an exception not the rule in ERA. Application of GLM for both redesign of algorithms and tuning on architecture available resources (supported by T-configurator at the element and architecture levels) maximize parallelism and minimise concurrency.
- Higher level rate of fault is expected for multi-core systems.	- Assumptions about higher rate of permanent hardware faults for the next generation of electronics are not correct. More details see, for example in Feynman lectures [Feynman].
- SEC/DED options proposed.	- Number of malfunctions caused externally by alpha particles and internally due to higher density of elements on the wafer does not let increase reliability using SEC/DED; most likely 16+ bit errors will take place in hardware. System software contribution to the malfunction of the system caused by support of dynamic parallelism and complex concurrency monitoring.
- Synchronization overheads should be reduced.	- Overloaded by task parallelism monitoring OS will be deadlocked. ERA proposes a re-configurability of available hardware and tasks starting at the algorithm compilation.
- Wide range of data types should be implemented: 1 bit (Boolean); 8 bits (Integer, ASCII); 16 bits (Integer, DSP fixed point, Unicode); 32 bits (Integer, Single-precision FP, Unicode);64 bits (Integer, Double-precision FP;128 bits (Integer, Quad-Precision FP; Large integer (>128 bits) (Crypto)	- Dual string approach for data structure in ERA where for each data element of the array a special descriptor defines data type (altogether, 2^{32} types). - This covers all possible data types user can dream of, or imagine. Efficiency of access to the proposed data structure is equal to the array access.
- New models should support proven styles of parallelism.	- Styles of parallelism are application specific and vary due to technology modification. ERA approach is proposing auto-tuning of existing programs into their maximum parallel forms.
- FPGA systems are future HW platforms for multi-core computing.	- FPGA technology at the element level and specially designed wafer with pre-fabricated configuration fabric of active processing element and passive elements enables to monitor architecture for the performance, power consumption or reliability.