The traditional ways of increasing hardware performance predicted by Moore’s Law have vanished. When uni-cores were the norm, hardware design was decoupled from the software stack, thanks to a well-defined Instruction Set Architecture. This simple interface allowed developers to design applications without much concern for the hardware, while hardware designers were able to exploit parallelism in superscalar processors. With the irruption of multi-cores and parallel applications, this approach no longer worked. As a result, the role of decoupling applications from the hardware was moved to the runtime system. Efficiently using the underlying hardware from this runtime without exposing its complexities to the application has been the target of research in the last years.

It is our position that the runtime has to drive the design of future multi-cores to overcome the restrictions in terms of power, memory, programmability and resilience that multi-cores have. In this talk, we introduce an approach towards a Runtime-Aware Architecture, a massively parallel architecture designed from the runtime’s perspective.