Circuit performance has increased tremendously over the past decades with the continuous scaling of CMOS technology. Typically hardware designers tend to employ conservative safety margins for timing closure. However, continuing with this approach could become very costly in the short term future. This is because the highly scaled CMOS devices would inevitably exhibit probabilistic behavior, meaning that we can no longer expect a uniform circuit performance. Covering all possible worst cases would become increasing difficult, expensive and result in large yield loss due to the variation.

While the standard approaches such as heavily pipeline the datapath can be used to boost the frequency, the overall latency will not tend to be reduced. As a result, this method will not be applicable to many embedded applications, which normally have strict latency requirements, or in any datapath containing feedback where C-slow retiming is inappropriate. Besides, the conservative timing margin will not be removed by these approaches. To tackle this problem, a large volume of current studies has demonstrated that relaxing the absolute accuracy requirement can provide the freedom to create designs with better performance or energy efficiency. Our previous work also demonstrated that allowing timing violations to happen would be of benefit in many applications, because the worst case only happens with a very small probability.