An unprecedented reengineering of the mobile networking architecture is required if the mobile industry is to remain competitive. Mobile data traffic is predicted to increase globally 11-fold from 2013 to 2018, and grow three times faster than fixed IP traffic. The growth in mobile traffic will not be homogenous: busy hour Internet traffic is expected to grow more rapidly than average Internet traffic. The uptake of Machine Type Communication (MTC) services will, in addition, result in spatially and temporally varying demand. Thus, a future mobile infrastructure will not only have to support a fast growing overall mobile data volume and a significantly increased number of connected mobile devices demanding significantly improved efficiencies, but it also has to flexibly adapt to dynamically fluctuating traffic demands and a broad range of potentially new requirements of future service portfolios.

To satisfy the above trends, future 5G networks will have to meet a wide array of diverse and extreme requirements. There will be the need for super fast and reliable connectivity with virtually zero latency for use cases such as remote control of robots, and support for billions of sensors and things. 5G will provide consistent and high quality connectivity for people and things, creating the perception of infinite capacity. Furthermore, 5G networks will combine revolutionary technologies with legacy existing mobile radio generations, as well as Wi-Fi, into a new system. Those technologies and layers will need to be managed as one.

So far, the 5G research community has been focused on novel physical layer technologies. However, it is essential to introduce a 5G End-to-End Architecture that allows for a cost- and time-efficient introduction of 5G, that integrates a variety of novel technologies rolled out in various stages, and that is sufficiently flexible to accommodate applications and services that are yet to be envisioned. For example, to allow 5G to support low latency, it is not sufficient to design a low latency RAN, since latency is an end-to-end factor and hence requires the joint design of RAN, Core and Transport Networks. Hence, the new E2E Architecture will have to be programmable, software driven and managed holistically to enable a diverse services in a profitable way. This requires novel ways of designing the mobile network architecture (radio access, core and transport network), operating the network, and offering interfaces.