# ITRS 2.0: Toward a Re-Framing of the Semiconductor Technology Roadmap

Juan-Antonio Carballo[3], Wei-Ting Jonas Chan[1], Paolo A. Gargini[4], Andrew B. Kahng[1,2] and Siddhartha Nath[2]

[1]ECE and [2]CSE Departments, UC San Diego, La Jolla, CA 92093
[3]Broadcom Corporation, San Jose, CA 95134
[4]Stanford University, Stanford, CA 94305
carballo@broadcom.com, wechan@ucsd.edu, pgargini@stanford.edu, abk@ucsd.edu, sinath@ucsd.edu

*Abstract*—**The International Technology Roadmap for Semiconductors** *(ITRS) has roadmapped technology requirements of the semiconductor industry over the past two decades. The roadmap identifies major challenges in advanced technology and leads the investment of research in a cost-effective way. Traditionally, the ITRS identifies major semiconductor IC products as* **drivers***; these set requirements for the state-of-the-art semiconductor technologies.* **High-performance microprocessor unit** *(MPU-HP) for servers and* **consumer portable system-on-chip** *(SOC-CP) for smartphones are two examples. Throughout the history of the ITRS, Moore's Law has been the main impetus for these drivers, continuously pushing the transistor density to scale at a rate of 2× per technology generation (aka "node"). However, as new requirements from applications such as data center, mobility, and context-aware computing emerge, the existing roadmapping methodology is unable to capture the entire evolution of the current semiconductor industry. Today, comprehending how key markets and applications drive the process, design and integration technology roadmap requires new system-level studies along with chip-level studies. In this paper, we extend the current ITRS roadmapping process with studies of key requirements from a system-level perspective, based on multiple generations of smartphones and microservers. We describe potential new system drivers and new metrics, and we refer to the new system-level framing of the roadmap as* **ITRS 2.0***.*

## I. INTRODUCTION

*The International Technology Roadmap for Semiconductors* (ITRS) [23] has for nearly two decades projected the next 15 years of technology requirements and potential solutions for the semiconductor industry over the past two decades. Historically, the ITRS uses metrics such as transistor density, number of cores, power, etc., to roadmap technology evolution of *integrated circuits* (ICs). These metrics are essentially driven by the physical-dimension scaling as predicted by Moore's Law. The current ITRS System Drivers Chapter [33] roadmaps key IC products that drive process/design technologies. However, new requirements from applications such as mobility, data centers, etc. require a new approach for system-level roadmapping because these applications drive technology roadmap for system-level metrics (e.g., the number of sensors, memory bandwidth, etc.). In this paper, we present a new roadmap for key systems that contain IC products and drive process/design/integration technologies. We refer to the roadmap with these new drivers and metrics as **ITRS 2.0**.

### Changes in the Semiconductor Industry Supply Chain

The 1980s and 1990s saw a semiconductor industry dominated by *integrated device manufacturers* (IDMs) [18]. During this period, the design architecture of the main driver in the ITRS, the *microprocessor unit* (MPU), was not application-driven. The standard components in the MPU, such as memories and microprocessors, scaled their densities and operating frequencies continuously to meet aggressive performance and cost requirements. Applications had to be designed based on these components. However, in the past ten years, fabless design houses have changed the industry landscape. Design teams have been building customized *system-on-chip* (SOC) and *system-in-package* (SIP) products to address specific application requirements instead of building standard components. As applications evolve, they drive further requirements for heterogeneous integration, outside system connectivity, etc. Up to the current 2013 edition, the ITRS roadmapping process (e.g., as seen in the System Drivers Chapter [33]) has not comprehended or roadmapped these system product-level requirements. Therefore, one of the crucial missions of the ITRS 2.0 is to connect emerging system product drivers and corresponding metrics into the roadmapping methodology. Two case studies illustrate how application domains induce new system drivers and metrics in the ITRS 2.0: (i) a *smartphone driver* driven by both fast-growing mobility and context-aware computing applications, and (ii) a *microserver driver* driven by a data center application. Though smartphones and microservers have existed for a number of years, it is in the context of requirements from key application domains that they emerge as system drivers in the ITRS 2.0 framework.

### Case Study 1: Smartphone

In recent years, mobile devices, notably smartphones, have shown significant expansion of their computing capabilities. This trend has been mentioned in the latest ITRS update [33] for the *consumer portable SOC* (SOC-CP) driver [2]. The current SOC-CP driver captures technology evolution and design challenges for a single IC. However, since smartphone systems are built with multiple heterogeneous ICs (e.g., logic, memory, *microelectromechanical systems* (MEMS) and *radio-frequency* (RF)), we must understand tradeoffs at the system level. For example, Chan et al. [2] note the trend of fast-growing 3D graphics in the SOC-CP driver, but the scope of their study does not include the connection between 3D graphics and display bandwidth. The ITRS 2.0 introduces a new smartphone driver to comprehend and roadmap metrics at a higher, system level for mobility applications.

### Case Study 2: Microserver

Recent studies of data centers (e.g., by Doller et al. [3]) suggest that *high-performance MPU* (MPU-HP) and *networking SOC* (SOC-NW) products are the main components in data centers. These products may be implemented either in a single chip or in a *multichip module* (MCM). The authors of [3] observe that an optimized data center achitecture cannot be achieved with a single chip as its key building block; rather a, co-optimization

of storage, interconnects and software is required. Since the raw data stored in data centers are usually sparse, pre-processing that is typically executed in traditional server cores is precluded, due to energy budget. Besides integrating power-efficient cores to be within an energy budget, data centers require high bandwidth and accessibility for local memories (mostly non-volatile memories) to execute data-intensive operations. Due to data center-specific optimizations and system-level design requirements such as high rack density and cooling capacity, the metrics of servers in data centers are different from those of server chips in existing products. The ITRS 2.0 introduces a new microserver driver to comprehend and roadmap metrics at a higher, system level for data center applications.

TABLE I: Two ITRS 2.0 system drivers discussed in this paper. Metrics driven by applications are shown in the right column.

| Application | ITRS 2.0 Driver | System Metrics |
|---|---|---|
| Data center | Microserver (new driver) | Computation performance, network throughput, space requirement, energy efficiency |
| Mobility | Smartphone (new driver) | Computation performance, multimedia performance, wireless bandwidth, thermal budget, battery time, BOM cost |

The remainder of this paper is organized as follows. Section II outlines the limitations of the current ITRS roadmap and motivates the introduction of new drivers based on application requirements. Section III describes the resulting new roadmapping methodology and Section IV provides example roadmaps of metrics for each of the new drivers. We conclude in Section V.

## II.   EVOLUTION OF SYSTEM DRIVERS

The current ITRS roadmapping process for system driver IC products (e.g., MPU-HP, SOC-CP, etc.) [33] does not consider application requirements, and focuses on roadmapping metrics at the IC level. New requirements from applications such as mobility, context-aware computing and the *Internet of Things* (IoT) [4] drive multiple technologies at the system level. A goal of ITRS 2.0 is to roadmap the impact of these applications on the semiconductor industry, hence our focus on system design instead of component ICs.

### A.  New Smartphone Driver

Figure 1, based on the Qualcomm Snapdragon™ family [34] of SOCs, illustrates the growth of features and degree of integration in recent *application processors*[1] (APs). Each new technology generation (aka "node"), which enables reduced computation power (e.g., new *instruction set architecture* (ISA), new devices, new low-power techniques) or introduction of new features (e.g., *graphic processing unit* (GPU) or 1080p video), brings an increased number of vertically-stacked bars in the plot. The figure shows that the

[1]Application processors essentially correspond to the SOC-CP products from the current ITRS roadmap (System Drivers Chapter) [33].

degree of integration after 2008 keeps increasing to meet the demands of (i) higher computation performance, (ii) faster wireless connections, and (iii) richer multimedia capabilities.

Figure 2 shows the number of ICs as another aspect of the evolution in a smartphone. By collecting the number of IC components from teardown analysis reports of smartphone devices from leading smartphone manufacturers, we observe that the number of IC components increases from 2002 to 2007, that is, during the first five years after the introduction of the smartphone. A possible explanation of this trend is that the required features rapidly increase when smartphones are differentiated from the previous feature phones. Figure 2 shows that the number of IC components used to construct a smartphone begins to decrease after 2007. The trends in the above two figures indicate that the demand for features, and the demand to reduce BOM cost, together drive the emergence of APs (i.e., the SOC-CP driver in the current ITRS) to most effectively utilize the transistors available from Moore's Law scaling.
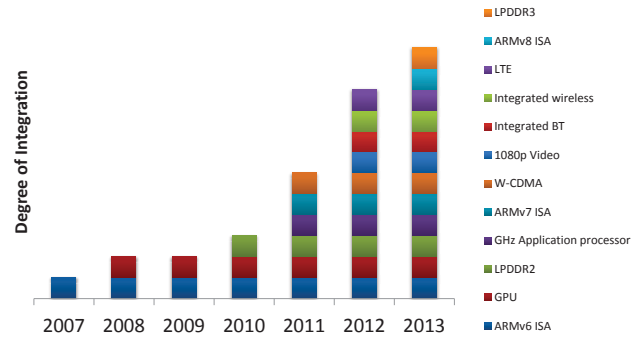


Fig. 1: Increasing degree of integration in mobile application processors (Qualcomm Snapdragon™ family).
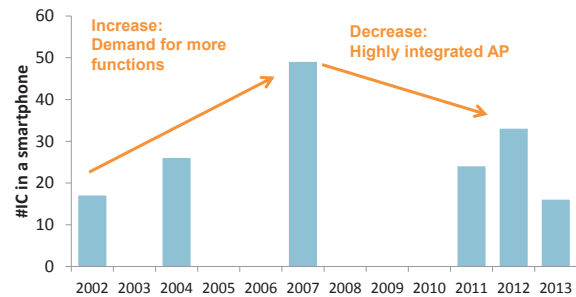


Fig. 2: Number of ICs in a specific smart phone product series.

### B.  New Microserver Driver

Another example of application-driven evolution is observed in servers. Traditionally, servers have driven the MPU-HP driver to achieve higher computation performance by scaling up the operating frequency, the number of cores and the sizes of cache memories. Since high-performance servers have dedicated cooling devices, the thermal limitations are much more relaxed compared to other system drivers in the ITRS. According to our studies, two factors lead to the introduction of the new microserver driver. First, cooling costs, which can reach over 35% of electricity costs

[5], continue to rise in server farms and data centers; this creates a need to reduce the number of cores and operating frequencies to limit this cost. Second, the demand for cloud computing requires a drastic reduction in communication latencies to meet an under-100$ms$ requirement [21], that is, data must be increasingly localized. Our initial studies suggest that the microserver driver addresses the cost issue by limiting the number of cores per rack unit and the latency issue by localizing user-specific search data [3]. The following discussion exapnds upon these issues.

The volume of information in data centers is anticipated to grow at a very high rate ($2\times$ every two years [15], or even faster [3]). When users search for specific information, latencies can be on the order of tens of milliseconds because data centers typically store information in a highly distributed manner. As data centers grow in size, communication latencies increase along with power consumption (e.g., $75MW$). To limit power and temperature of data centers, companies are forced to invest huge amounts of money to establish and maintain power plants adjacent to data centers, and to construct data centers in geographies with "natural refrigeration". There is a limit to such investment in power plants and cooling. The microserver is possibly a new solution that can lower costs and power in data centers.

Microservers must also process huge numbers of queries with reduced latency. Therefore, this driver must maximize the number of cores in a rack unit subject to power and thermal constraints. Form factor, energy-efficiency, and networking throughput are important metrics for this driver. Table II compares the specifications of a mainstream 1U server [16] and a microserver cartridge [20]. Since the thermal limit of a microserver is significantly stricter than that of a conventional 1U server, the available computation performance and networking bandwidth are quite constrained. As a consequence, demand for reduced form factor and system design effort drives the integration of the MPU and the chipset. Compared to a 1U server (MPU-HP), a microserver has a higher degree of integration as it includes on-chip Ethernet and peripheral hubs. Recent microserver MPUs [24] integrate application-specific accelerators to improve energy efficiency. Hence, the ITRS 2.0 roadmapping process must consider application requirements for system-level integration.

Another motivation for the evolution of microservers as a system driver is that data centers can be fragmented to host user-specific queries. Historically, it has been difficult to guess the kind of information that would be useful to users, and data centers have therefore been designed to host almost everything. As more information is collected on users' typical searches, it may be possible to restructure data centers into smaller units optimized for "localized usage". This approach would further reduce search time and lead to smaller, more localized, data centers with reduced power requirements and costs.

### III.   ITRS 2.0 ROADMAPPING METHODOLOGY

We now describe an initial methodology to roadmap system drivers in the ITRS 2.0. The flow is summarized in Figure 3. First, we collect details of IC components from the following data sources: (i) published data from web searches [10] [28] [34] [36], (ii) specification documents, datasheets and whitepapers from IC companies [9] [16] [20] [24], (iii) teardown reports [29], and (iv) high-level comments from industry collaborators. Second,

TABLE II: Comparison of system specifications between a conventional 1U server and a microserver.

|  | 1U Server | Microserver |
|---|---|---|
| System TDP ($W$) | 500 | 100 |
| TDP per MPU ($W$) | 150 | 12 |
| Memory ($GB$) | 768 | 8 |
| Memory BW ($GB/s$) | 100.0 | 10.7 |
| Ethernet BW ($GB/s$) | 20 | 2 |
| Main system chips (MPU and chipset) | 3 | 1 |

we develop function categories for IC components by classifying IC components that provide similar functionalities into the same category. Based on the classification, we create abstract block diagrams as system models. We also analyze the components and predict how metrics such as maximum operating frequency, die area, number of antennas, number of sensors, etc. evolve during the course of the roadmap. Finally, we produce a roadmap for system-level metrics based on the projected metrics and the abstract block diagrams. We believe that these steps can provide us with insights into how drivers evolve from application requirements. The remaining of this section illustrates our application of this methodology to smartphones and microservers in the ITRS 2.0 roadmap.
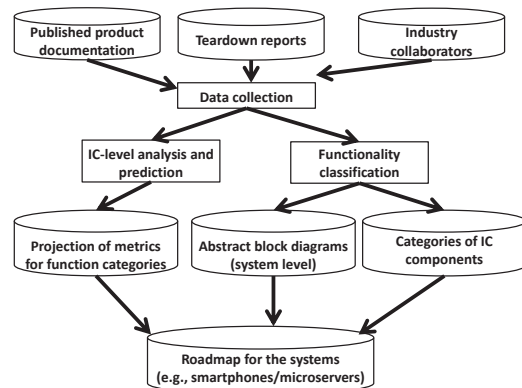


Fig. 3: Flow of data collection, analysis, and metric projection in the ITRS 2.0 roadmapping methodology.

### A. Functional Components in Smartphones

To roadmap smartphones, we conduct studies of IC components in state-of-the-art smartphones, which are at the leading edge for metrics such as the number of frequency bands used for the radio, multimedia capabilities, storage capacities, etc. We then analyze the number of ICs in terms of function categories in smartphones and determine the scaling trajectories. The functional components in a 2012 smartphone [29] are summarized in Table III. We expand upon these components in smartphones as follows. (i) **APs** are used to implement key functionalities such as system resource management, scheduling, and file system. *Baseband processors* (**BBs**) manage the wireless connections between carrier base stations and smartphones. In earlier smartphones, multimedia features are implemented by coprocessors. We also include

141

coprocessors in this category. (ii) **Memory** refers to DRAM and flash memories used for storage. (iii) *Radio-frequency* (**RF**) refers to IC components which transmit and receive radio-frequency signals. This category includes *power amplifiers*, *transceivers*, *antenna switches*, *RF filters*, etc. (iv) **Sensors** refer to any IC component that can capture images or sense surrounding temperature, lighting, phone postures, and other user activities and environmental changes. (v) **Display/input devices** refer to display controllers and touchscreen controllers. (vi) **Power/Analog** refer to power management ICs, LED controllers and audio amplifiers. (vii) **Connections** include both wireless and wired connections to services or devices other than cellular base stations, such as Bluetooth, WiFi, USB and GPS.

TABLE III: Functional components of a 2012 smartphone [29].

| Subsystem | Components | Functions |
|---|---|---|
| AP | Application processor | Multimedia, resource management |
| BB | Baseband | Communication protocols |
| Memory | Mobile DDR2 SDRAM memory | |
| | MLC NAND flash memory | Data storage |
| | Memory controller | |
| RF | RF power amplifier | Radio signal processing |
| | Power amplifier controller | Up/down frequency conversion |
| | WiFi 802.11n / Bluetooth (BT) / FM radio | Radio band management |
| | GSM, CDMA, W-CDMA, GPS transceivers | |
| | SP10T antenna switch | |
| | GPS LNA | |
| Sensor | LED flash driver | Image capturing |
| | 5 MP CMOS image sensor | |
| | Accelerometer processor | Position / orientation detection |
| | MEMS sensor | |
| | Ambient light / proximity sensor | Environment sensing |
| Display/input device | Capacitive touchscreen controller | Touch control |
| | TFT-LCD display driver | Display |
| Power/analog | Power management IC (PMIC) | Power mode management |
| | Audio CODEC / amplifier | Audio output |
| Connections | Interface circuits or IO controllers | Connections between devices |

### B. Abstract Block Diagrams of Smartphones for Analyses

We now describe construction of abstract block diagrams consistent with the IC function categories discussed in Section III-A, and using data collected from, e.g., teardown reports [29] of leading-edge smartphones. To contrast the differences between early smartphones and recent ones, we choose one smartphone from 2002. We classify the IC components used to build smartphones into the following function categories: *AP/BB+Multimedia*, *Connections*, *RF*, *Sensor*, *Power/analog*, *Display/input devices* and *Memory*. The 2002 smartphone contains the following components. (i) Several discrete IC components (VCO, TCXO, etc.) are used to build RF functionalities. (ii) The baseband processor is used to control RF circuits with an analog interface IC. (iii) An additional image processor is used to provide multimedia capabilities. (iv) Bluetooth modules are implemented using multiple chips.

Based on analysis of teardown reports, we exhibit two alternatives, *Alt-1* and *Alt-2*, for future smartphones in Figures 4(b) and 4(c). After years of evolution, following are the integrations we observe in the smartphone design. (i) All the multimedia functions are integrated into a single application processor, which can also provide baseband functionalities. (ii) Although RF supports multimode standards, the components are integrated to reduce the number of IC components. (iii) Bluetooth and WiFi are integrated into a single IC. (iv) Multiple sensors are integrated. (v) Flash memory capacity grows from tens of *MB* to *GB*, and memory chips are integrated along with the memory controller

into a single module. (vi) Touchscreen is the main input device, which requires independent controllers. In short, the integration of multiple heterogeneous functions in smartphones has been growing over the years. It is reasonable to project that smartphones will use the organization in Alt-1 up to ∼2019, and the organization in Alt-2 from ∼2020 onwards.

We summarize the Alt-1 and Alt-2 abstract smartphone block diagrams beyond 2013, as follows. (i) In each function category of Alt-1 and Alt-2, features as well as the supported industrial standards increase, such as multimedia and 3D graphics in the AP/BB category and multimode RF circuits in the RF category. (ii) Although the required features are increasing, the 2013 smartphones use a single AP in both Alt-1 and Alt-2 to avoid two additional coprocessors as seen in the 2002 smartphone. (iii) It is common to store multiple movies in smartphones which consequently increases the size of non-volatile storage. As a result, standard components such as DRAMs and flash memories need to scale up their capacities. Due to performance considerations, a discrete flash memory controller is added to the memory module in both Alt-1 and Alt-2. (iv) Both Alt-1 and Alt-2 integrate a richer set of sensors (e.g., image sensors and environment sensors) to monitor environmental events compared to that of the 2002 smartphone. In addition, a touchscreen controller is integrated to replace the conventional keyboard. (v) We expect aggressive power management with multiple sensors (e.g., low-power sensor hub of Alt-1 in Figure 4(b) or low-power (LP) cores in Alt-2 in Figure 4(c)). For MEMS-based sensors, we expect better integration, e.g., six axes (gyro × accelerometer) in a package, for both Alt-1 and Alt-2. (vi) We expect more reconfigurability in the baseband modem design from 2020 onwards, so the *software-defined radio* (SDR) is introduced in Alt-2. (vii) The overall application processor design introduces more heterogeneous cores in Alt-2, which will address power management issues with low-power cores (e.g., big.LITTLE architecture from ARM [13]).

### C. Functional Components in Microservers

We similarly extract main functional components in microservers from a latest product line of HP [20]. In additional to these basic functional components, we add several advanced functional components such as *embedded DRAM* (eDRAM) or optical interconnects to address the technology evolution. The functional components of a 2012 microserver [20] are summarized in Table IV. We expand upon these components in microservers as follows. (i) **Processor** components include different MPUs to process application tasks. Besides traditional high-performance cores, this component includes low-power cores to meet strict power constraints of microservers as well as accelerators to process data-related tasks. (ii) **Memory** includes conventional high-performance DRAM and eDRAM. Microservers achieve greater power efficiency and performance by integrating eDRAM close to the processors. (iii) **On-board switching** components in microservers include interfaces to manage data transfers, such as serial links (e.g., PCI Express) and optical interconnects to provide high networking data rates. (iv) **Networking** components include gigabit Ethernet and fiber optic networks [7]. (v) **Peripherals** include components to monitor system health, and remotely maintain and manage IC components. (vi) **Storage** includes local storage (e.g., hard drives or solid state drives) as well as networked storage.
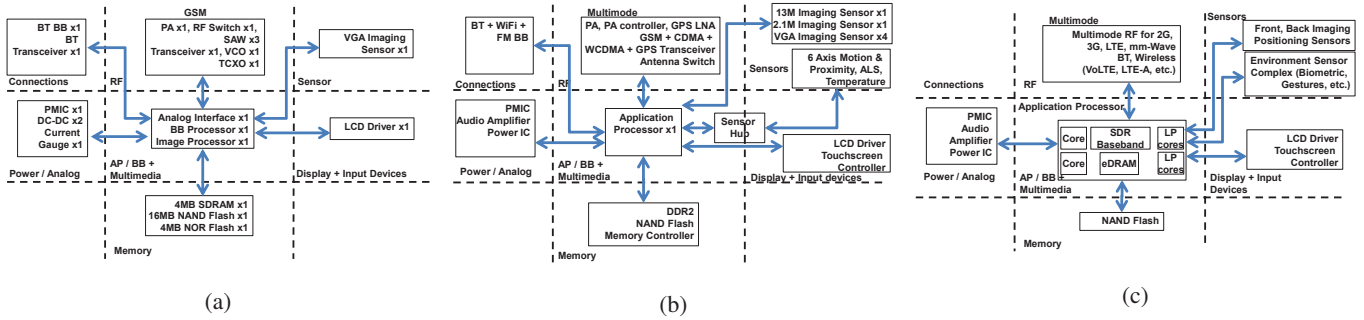
Fig. 4: Abstract block diagrams of smartphones: (a) 2002, (b) Alt-1 projection, and (c) Alt-2 projection.
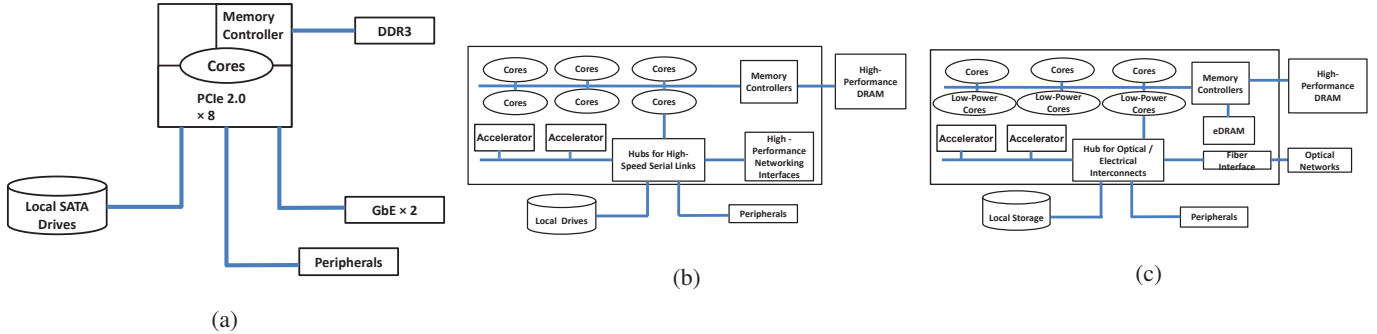


Fig. 5: Abstract block diagrams of microservers: (a) 2012, (b) Alt-1 projection, and (c) Alt-2 projection.

TABLE IV: Functional components of a 2012 microserver [20] and beyond. (Asterisks denote emerging technologies.)

| Subsystem | Component | Functions |
|---|---|---|
| Processor | MPU core | General computation, encryption, memory control, on-chip communication |
| | *MPU core (low power) | |
| | Memory controller | |
| | *Accelerator | |
| | On-chip fabric | |
| Memory | DRAM | High bandwidth local data storage |
| | *eDRAM | |
| On-board switching | Serial link controller | Interconnection among local devices |
| | *Local optical link | |
| | Interconnect hub | |
| Networking | Ethernet controller | Interconnection among servers |
| | *Optical network controller | |
| Peripheral | System monitor | System health monitor, failure analysis |
| | Management module | |
| Storage | Hard drive | Bulk data storage |
| | Flash memory | |
| | Storage/RAID controller | |

### D. Abstract Block Diagrams of Microservers for Analyses

Figure 5(a) analyzes the structure of microservers in existing 2012 products and projects the block diagram evolution with two alternatives (*Alt-1* in Figure 5(b) and *Alt-2* in Figure 5(c)). To address the power density challenge, microservers will integrate the following features in both Alt-1 and Alt-2: (i) low-power light cores, (ii) integrated peripherals, (iii) application-specific accelerators, and (iv) high throughput inter-IP communication fabrics. To address system throughput requirements, microservers will integrate products that provide large networking bandwidth in both Alt-1 and Alt-2.

The Alt-1 organization is similar to that of microservers in 2012, with some architectural improvements in MPUs and peripherals. The timeline for such an organization might be up to ~2019. Alt-2 integrates optical interconnects and heterogeneous cores. A reasonable timeline for such an organization is from ~2020 onwards.

### IV. ROADMAPS FOR THE NEW ITRS 2.0 DRIVERS

From the Section III methodology and abstract block diagrams, initial roadmapping may be performed. We now describe prototype roadmaps for the smartphone and microserver drivers.

#### A. Roadmap for Smartphones

Key trends are summarized in Table V. We categorize the metrics into two types: (i) input metrics, which are defined as the metrics that can be directly obtained from product specifications, and (ii) output metrics, which are derived from the input metrics and the understanding of system architectures.

**Board area.** The small form factor of smartphones is one of the challenges to system designers. Width and height scaling of the smartphone board to accommodate larger displays, variety of sensors, antennas, memories and cores, ESD, etc. exposes an eventual gap of $\sim46cm^2$ in board area as shown in Figure 6(a). This gap in area scaling can (i) prevent integration of new functionalities such as MEMS-based environment sensors, (ii) cause thermal issues due to closely-packed components on the board, and (iii) increase the number of layers of the *printed circuit board* (PCB) and thereby increase the board cost. These challenges can be partially overcome by more pervasive use of 3D integration. This gap also suggests that innovations are required in board design, assembly and packaging, and high-density heterogeneous

TABLE V: Summary of scaling trends of smartphones.

| | Metrics | Metric Evolution |
|---|---|---|
| **Input Metrics** | #AP cores | $1.0 \times 1.14^{(Year-2007)}$ |
| | #GPU cores | $2.0 \times 1.26^{(Year-2007)}$ |
| | Max frequency $(GHz)$ | $2.3 \times 1.04^{(Year-2011)}$ |
| | #MPixels of display | $0.7 \times 1.27^{(Year-2013)}$ |
| | Memory BW $(Gb/s)$ | $16000 \times 1.19^{(Year-2011)}$ |
| | #Sensors | $7 + (Year - 2009) \times 0.85$ |
| | #Antennas | $7 + (Year - 2009) \times 0.4$ |
| | #ICs | $10 - (Year - 2009) \times 0.15$ |
| | Cellular data rate $(MB/s)$ | $5.25MB/s$ until 2022, $21.6MB/s$ afterwards |
| | WiFi data rate $(Mb/s)$ | $75Mb/s$ until 2015, $850Mb/s$ afterwards |
| **Output Metrics** | Board area $(cm^2)$ | See Figure 6 |
| | System power $(mW)$ | See Figure 6 |

3D integration of RF/AMS/MEMS and digital components to fit within a desired area budget of around $60cm^2$.

**System power.** The system-level power projections of the two alternatives in smartphone organization (Alt-1 and Alt-2) are respectively shown in Figures 6(b) and 6(c). We assume the following in making these projections.

- The baseline system power values in 2013 shown in Figures 6(b) and 6(c) are based on measurements from [27].

- For connection, either WiFi or cellular is fully enabled. These two connections are mutually exclusive, i.e., one will consume only idle power when the other is enabled.

- Main processor cores and GPU cores in application processors are active. GPU power is assumed to be twice the CPU power after calibration with data in [6].

- The activity factor of flash memory access is fixed at 10%.

- We calculate the total power as the sum of wireless idle power and cellular active power (or, wireless active power and cellular idle power), AP power, GPU power, display power, and memory power.

- Idle power values are fixed in both Alt-1 and Alt-2.

The baseline of power growth of each IC component in smartphones is the same as the 7% per year in current SOC-CP power projections [2] [33]. We consider two alternatives to roadmap system power, Alt-1 (4% growth in power per year) and Alt-2 (5% growth in power per year) in Figures 6(b) and 6(c), respectively. Alt-1 indicates a $3.3W$ power gap and Alt-2 indicates a $4.5W$ power gap at the 15-year horizon. The gap of board-level power will result in the following design challenges. (i) Since battery capacity increases slowly, the power gap (as high as $4.5W$) may significantly challenge battery life of smartphones. (ii) Higher power consumption will generate more heat in smartphones and will make thermal design more challenging within a small form factor. We expect that more aggressive low-power design techniques, such as power-efficient cores, will be deployed to mitigate the power challenge. 3D integration technologies may also contribute to power reduction by reducing the loading of IO circuits.

**Number of pixels.** Figure 7(a) shows the scaling of the number of pixels in smartphone displays. Display pixels are driven by high-

definition standards (e.g., 720p, 1080p, 4K, etc.). We observe that the number of bits per pixel and refresh rates do not scale very significantly over the years (e.g., refresh rate is limited to $120Hz$ for 2D, and $240Hz$ for 3D). Increase in the display size increases the memory bandwidth requirement as shown in Figure 7(b). By 2029, ultra HD resolutions of $7680 \times 4320$ could potentially increase memory BW requirements to $144Gb/s$ [39].[2]
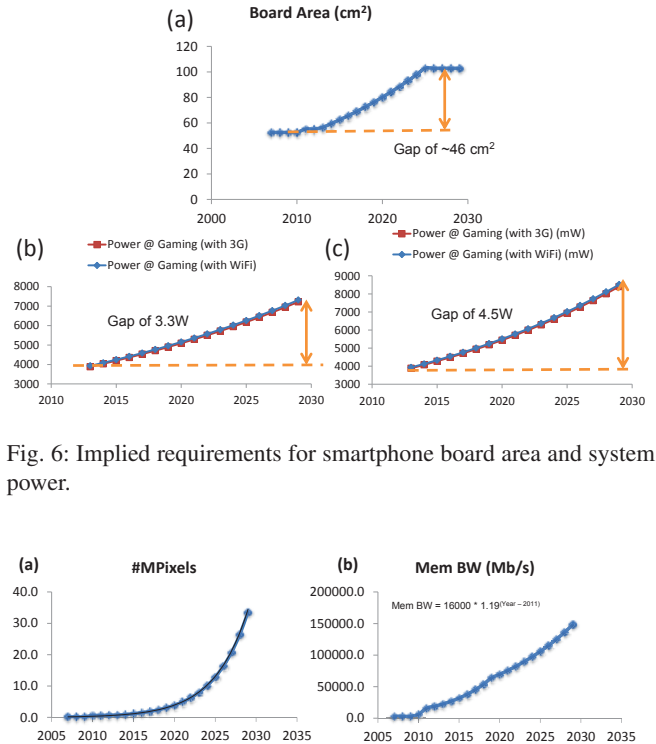


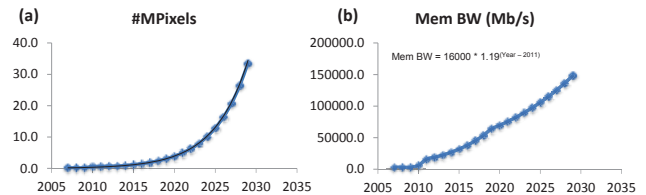Fig. 6: Implied requirements for smartphone board area and system power.



Fig. 7: Scaling of display size and memory bandwidth.

**Number of IC components.** From the scaling plot in Figure 8(a), we observe rapid growth in the number of ICs from 2007 to 2013, driven by escalating demand for smartphone features. From 2013, we observe that there are two driving forces for the number of ICs which compete with each other through the roadmap: (i) integration technology decreases the number of ICs used in smartphones, and thus reduces the number of components and the complexity of system integration, PCB board processing, etc.; and (ii) demands for new functionalities increase the number of ICs in order to deliver new features in smartphones. Further, unavailable integration or low-power design technologies may also prevent integrations of certain functionalities; this can cause some blocks to be dis-integrated from their original host ICs. Contention between these two driving forces implies that the scaling of number of ICs in smartphones will not be monotonic. We expect that the trend will converge to less than 10 ICs at the end of the roadmap, but will not decrease further due to fundamental limitations to integration of heterogeneous ICs (e.g., logic, MEMS, RF, etc.). From our studies of smartphones, we may discern an example of integration

---

[2]We note that this memory BW projection for 2029 may well be conservative: recent 5G technology roadmaps suggest a BW requirement of $\sim 100Gb/s$ in 2020 (e.g., [8]).

of ICs for RF circuits, and an example of dis-integration of ICs for sensors.
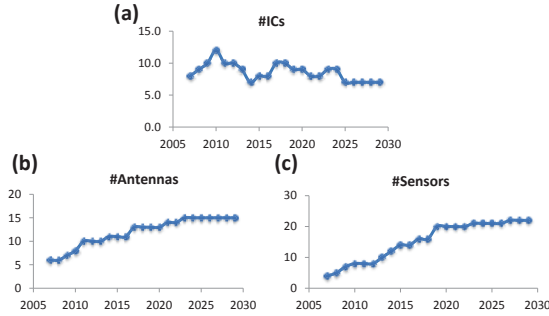


Fig. 8: Scaling of numbers of (a) ICs, (b) antennas, and (c) sensors by category in a smartphone from 2007 to 2013.

The design complexity of RF circuits of smartphones can be quantified by the scaling of the number of antennas in smartphones as shown in Figure 8(b). The scaling is driven by the implementation of WiFi, BT, cellular, and *mm*-wave applications. However, IC designers and system integration companies will necessarily seek enhanced integration technologies of RF circuits to maintain reasonable system cost. For example, RF front-end circuits are conventionally implemented using GaAs technology as it provides better electrical properties, such as carrier mobility. The left plot in Figure 9 shows a traditional system design, wherein RF circuits (PA, antenna switches and RF filters) and controllers (digital) are distributed across different IC components. ICs with RF modules require relatively larger board area and more complicated board-level signal integrity considerations. The right plot shows an integrated solution enabled by silicon RF technology and 3D integration; these technologies simplify system-level design efforts at the cost of losing device electrical characteristic advantages of GaAs technology.
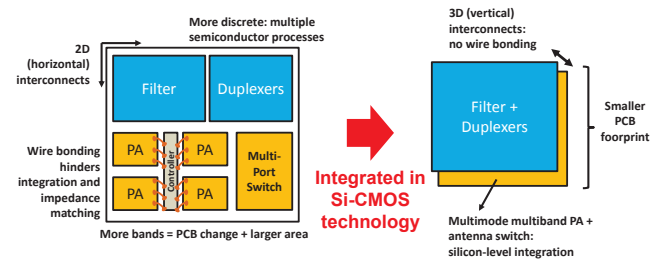


Fig. 9: Example of integration: Qualcomm CMOS RF front-end circuit, reproduced from [1].

The number of environment sensors increases linearly as shown in Figure 8(c). Since these sensors monitor both environment changes and user behavior to meet application requirements (e.g., IoT), the sensor data must be continuously collected. In recent Apple smartphones, the data collection and analysis tasks are assigned to a coprocessor (Apple M7 [12]), referred to as the "sensor hub" in the system. We contrast the technologies and chip specifications of two smartphones in Table VI. The data indicates that even though the Apple A7 processor die has more transistor capacity in which to integrate new features, the sensor hub is still moved out to a coprocessor manufactured in an older

technology and slow operating frequency. A possible reason for the dis-integration is that the always-on tasks for the sensor hub cannot be hosted in the AP with high energy efficiency. This suggests that better low-power design techniques are a minimum requirement for integration of sensor hub/always-on components within an AP.

TABLE VI: A dis-integration example found in the Apple A7 platform. Although the transistor capacity of the A7 die has increased over the A6 die, the sensor hub (M7) is still moved out of the A7 for energy-efficiency reasons.

| | iPhone 5 | coprocessors in iPhone 5s | |
| | **A6** | **A7** | **M7** |
|---|---|---|---|
| Frequency (*GHz*) | 1.30 | 1.40 | 0.15 |
| Node (*nm*) | 32 | 28 | 90 |
| Area (*mm$^2$*) | 97 | 102 | N/A |
| Normalized Tx capacity | 1 | 1.37 | N/A |

### B. Roadmap for Microservers

To meet application requirements driven by the data center, microservers have the following design targets: (i) good energy efficiency to reduce facility operation overheads such as cooling or power supplies, (ii) flexible network interfaces to build efficient clusters, and (iii) small form factor to integrate plenty of computing resources, including processors and memories, in limited spaces. Based on these design criteria, we project the metrics summarized in Table VII to study the evolution of metrics driven by the microserver. Most of the jobs on microservers are batch-processed, which drives growth in the number of low-power cores. The growth is constrained by power and cost. The frequency roadmap is the same as that seen in the existing MPU-CP roadmap, and is driven by real-time jobs (e.g., video streaming, maps, visualization). The roadmap is limited by power, which grows to around 90*W* in 2029, again driving low-power circuit and design methodologies. DRAM capacity per rack unit (1U) is driven by the number of cores [22] in a rack, application locality needs, and multithreading or multiprocessing. DRAM bandwidth is driven by faster clocks [25], the number of cores and smaller latency requirements for high-priority tasks (e.g., advertisement and map services). Off-MPU bandwidth is driven by both batch and real-time applications (e.g., queries and video streaming).

According to Moore's Law, the scaling of transistor count is at least 1.26× per year, given a three-year technology node cycle. We observe some outliers in the scaling of certain metrics relative to this growth rate. Growth of DRAM capacity per rack unit is much higher than this threshold, which indicates an aggressive memory requirement due to high-volume data. Scaling of the number of cores per rack unit and maximum clock frequency is slower than 1.26× per year, which indicates that the designs invest more resources on peripherals, interconnects, and memory interfaces instead of the number of cores.

## V. CONCLUSIONS

This paper describes the limitations of the current ITRS roadmap with respect to providing a system-level roadmap

TABLE VII: Summary of scaling trends of microservers.

| Metrics | | Metric Evolution |
|---|---|---|
| **Metrics** | #MPU cores<br>per rack unit | $16 \times 1.19^{(\text{Year}-2013)}$; Year $\leq 2019$<br>$45 \times 1.12^{(\text{Year}-2019)}$; Year $\geq 2019$ |
| | Max frequency ($GHz$) | $3.46 \times 1.04^{(\text{Year}-2013)}$ |
| | DRAM capacity ($GB$)<br>per rack unit | $128 \times 1.58^{(\text{Year}-2013)}$ |
| | DRAM BW ($GB/s$) | $51.2 \times 1.26^{(\text{Year}-2013)}$ |
| | Off-MPU BW ($GB/s$) | $64 \times 1.28^{(\text{Year}-2013)}$; Year $\leq 2019$<br>$285 \times 1.18^{(\text{Year}-2019)}$; Year $\geq 2019$ |
| | MPU frequency $\times$ #Cores<br>($GHz$) per rack unit | $55 \times 1.24^{(\text{Year}-2013)}$; Year $\leq 2019$<br>$45 \times 1.16^{(\text{Year}-2019)}$; Year $\geq 2019$ |

for "new" drivers that are driven by new requirements from applications such as mobility, context-aware computing, data center, etc. We sketch an initial ITRS 2.0 system roadmapping methodology using the smartphone and the microserver as exemplary system drivers. The methodology projects metrics and abstract block diagrams that describe (future) organization of these drivers. We identify metrics for each driver that we believe will drive various technology roadmaps for devices, interconnects, assembly and packaging, heterogeneous integration and outside system connectivity. It should be noted that solutions to the design challenges that we identify (e.g., power gap, board area gap, etc.) have not been completely explored in the current development that we present in this paper. Our ongoing work includes (i) identifying solutions to key gaps (board area, power, etc.) such as 3D integration, optimal split of logic die, optimal SIP or 3D integration, etc.; and (ii) specifying phases of solutions such as research, development, deployment, and dates when these solutions will be available using color codes as in the current ITRS roadmap.

### REFERENCES

[1] P. Carson and S. Brown, "Less is More: The New Mobile RF Front-End", http://www.microwavejournal.com/articles/20008-less-is-more-the-new-mobile-rf-front-end

[2] W.-T. J. Chan, A. B. Kahng, S. Nath and I. Yamamoto, "The ITRS MPU and SOC System Drivers: Calibration and Implications for Design-Based Equivalent Scaling in the Roadmap", *Proc. ICCD*, 2014, to appear.

[3] E. Doller, A. Akel, J. Wang, K. Curewitz and S. Eilert, "DataCenter 2020: Near-memory Acceleration for Data-oriented Applications", *Proc. Symposium on VLSI Circuits*, 2014, pp. 1-4.

[4] J. Gubbi, R. Buyya, S. Marusic and M. Palaniswami, "Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions", *Future Generation Computer Systems* (29) (2013), pp. 1645-1660.

[5] V. Kontorinis, L. E. Zhang, B. Aksanli, J. Sampson, H. Homayoun, E. Pettis, D. M. Tullsen and T. S. Rosing, "Managing Distributed UPS Energy for Effective Power Capping in Data Centers", *Proc. ISCA*, 2012, pp. 488-499.

[6] A. Pathania, Q. Jiao, A. Prakash and T. Mitra, "Integrated CPU-GPU Power Management for 3D Mobile Games", *Proc. DAC*, pp. 1-6.

[7] H. M. G. Wassel, D. Dai, M. Tiwari, J. K. Valamehr, L. Theogarajan, J. Dionne, F. T. Chong and T. Sherwood, "Opportunities and Challenges of using Plasmonic Components in Nanophotonic Architectures", *Trans. IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 2(2) (2012), pp. 154-168.

[8] "5G: A Technology Vision", *white paper*, Huawei Technologies Co. Ltd., 2013.

[9] http://www.apple.com/iphone-5s/features

[10] http://en.wikipedia.org/wiki/IPhone#Sensors

[11] http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/inside-the-iphone-5s

[12] http://en.wikipedia.org/wiki/Apple_M7

[13] http://www.arm.com/products/processors/technologies/biglittleprocessing.php

[14] *Personal communication*, Chipworks, August-September 2013.

[15] https://intel.activeevents.com/sz14/connect/fileDownload/session/A7DDB8CBFF9792C2E76059DB267A1100/SZ14_SSDS003_100_ENGf.pdf

[16] http://www.dell.com/downloads/global/products/pedge/en/Dell_PowerEdge_R620_Spec_Sheet.pdf

[17] http://www.infotech.com/download/50246

[18] Fabless versus IDM Company IC Sales from IC Insights, http://itersnews.com/?p=31514

[19] Google Trend, http://www.google.com/trends

[20] HP Moonshot System, http://h17007.www1.hp.com/us/en/enterprise/servers/products/moonshot/index.aspx

[21] http://www.ecoustics.com/products/hp-z800-z600-z400-workstations

[22] https://www.power.org/wp-content/uploads/2012/10/15.-IBM-DB2_lui-v5.pdf

[23] *International Technology Roadmap for Semiconductors*, http://public.itrs.net

[24] Intel Avoton processor, http://ark.intel.com/products/codename/54859/Avoton

[25] http://www.theregister.co.uk/2013/09/04/intel_avoton_rangeley_atom_c2000

[26] http://en.wikipedia.org/wiki/List_of_device_bit_rates

[27] http://mobiledevices.kom.aau.dk/research/energy_measurements_on_mobile_phones/results

[28] Nokia product list on Wikipedia, http://en.wikipedia.org/wiki/List_of_Nokia_products

[29] *Personal communication*, TechInsights, http://www.techinsights.com

[30] System Drivers Chapter in ITRS 2001 Update, http://www.itrs.net/Links/2001ITRS/Home.htm

[31] System Drivers Chapter in ITRS 2005 Update, http://www.itrs.net/Links/2005ITRS/Home.htm

[32] System Drivers Chapter in ITRS 2007 Update, http://www.itrs.net/Links/2007ITRS/Home.htm

[33] System Drivers Chapter in ITRS 2013 Update (to appear), http://www.itrs.net/Links/2013ITRS/Home.htm

[34] http://en.wikipedia.org/w/index.php?title=Snapdragon_(system_on_chip)

[35] http://http://www.invensense.com/mems/gyro/mpu6500.html

[36] http://www.techinsights.com/DeviceProfileSF_AustinParts.aspx?TeardownId=1683

[37] http://www.anandtech.com/show/7126/the-arm-diaries-part-2-understanding-the-cortex-a12

[38] http://www.anandtech.com/show/6747/htc-one-review

[39] http://http://en.wikipedia.org/wiki/Ultra_high_definition_television