1

The Future of CMOS: More Moore or a New Disruptive Technology?

Nazek El-Atab and Muhammad M. Hussain

King Abdullah University of Science and Technology, Integrated Nanotechnology Lab, Thuwal, 4700, Saudi Arabia

For more than four decades, Moore's law has been driving the semiconductor industry where the number of transistors per chip roughly doubles every 18–24 months at a constant cost. Transistors have been relentlessly evolving from the first Ge transistor invented at Bell Labs in 1947 to planar Si metal-oxide semiconductor field-effect transistor (MOSFET), then to strained SiGe source/drain (S/D) in the 90- and 65-nm technology nodes and high-κ/metal gate stack introduced at the 45- and 32-nm nodes, then to the current 3D transistors (Fin field-effect transistors (FinFETs)) introduced at the 22-nm node in 2011 (Figure 1.1). In extremely scaled transistors, the parasitic and contact resistances greatly deteriorate the drive current and degrade the circuit speed. Thus, miniaturization of devices so far has been possible due to changes in dielectric, S/D, and contacts materials/processes, and innovations in lithography processes, in addition to changes in the device architecture [1, 2].

The gate length of current transistors has been scaled down to 14 nm and below, with over 10^9 transistors in state-of-the-art microprocessors. Yet, the clock speed is limited to 3–4 GHz due to thermal constraints, and further scaling down the device dimensions is becoming extremely difficult due to lithography challenges. In addition, further scaling down the complementary metal-oxide semiconductor (CMOS) technology is leading to larger interconnect delay and higher power density [3]. The complexity of physical design is also increasing with higher density of devices. So, what is next?

A promising More-than-Moore technology is the 3D integrated circuits (ICs) which can improve the performance and reduce the intra-core wire length, and thereby enable high transfer bandwidth with reduced latencies and power consumption, while maintaining compact packing densities [4]. Alternative technologies that could be promising for new hardware accelerators include resistive computing, neuromorphic computing, and quantum computing.

Resistive computing could lead to non–von Neumann (VN) computing and enforce reconfigurable and data-centric paradigms due to its massive parallelism and low power consumption [5]. Moreover, humans can easily outperform current high-performance computers in tasks like auditory and pattern recognition

Advanced Nanoelectronics: Post-Silicon Materials and Devices,
First Edition. Edited by Muhammad Mustafa Hussain.

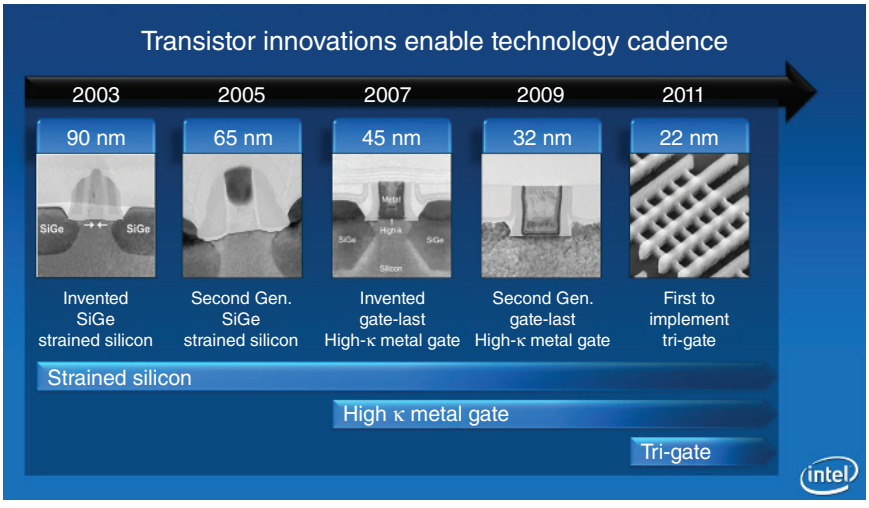

Figure 1.1 Intel innovation in process technology for the past decade. Source: www.intel.in.

and sensory motor control. Thus, neuromorphic computing can be promising for emulating such tasks due to its energy and space efficiency in artificial neural network applications [6]. Quantum computing can solve tasks that are impossible by classical computers, with potential applications in encryptions and cryptography, quantum search, and a number of specific computing applications [7].

In this chapter, four main technologies are discussed: FinFET, 3D IC, neuromorphic computing, and quantum computing. The state-of-the-art findings and current industrial state in these fields are presented; in addition, the challenges and limitations facing these technologies are discussed.

1.1 FinFET Technology

Over the past four decades, the continuous scaling of planar MOSFETs has provided an improved performance and higher transistor density. However, further scaling down planar transistors in the nanometer regime is very difficult to achieve due to the severe increase in the leakage current I_{off}. In fact, as the channel length in planar MOSFETs is reduced, the drain potential starts to affect the electrostatics in the channel and, consequently, the gate starts to lose control over the channel, which leads to increased leakage current between the drain and source. A higher gate-channel capacitance can relieve this problem using thinner and high-κ gate oxides; however, the thickness of the gate oxide is fundamentally restricted by the increased gate leakage and the gate-induced-drain leakage effect [8–10].

An alternative to planar MOSFETs is the multiple-gate FETs (MuGFETs) which demonstrate better electrostatics and better screening of the drain from the gate due to the additional gates covering the channel [11–14]. As a result, MuGFETs show better performance in terms of subthreshold slope, threshold voltage (V_t) roll-off, and drain-induced barrier lowering (DIBL). Another alternative to planar

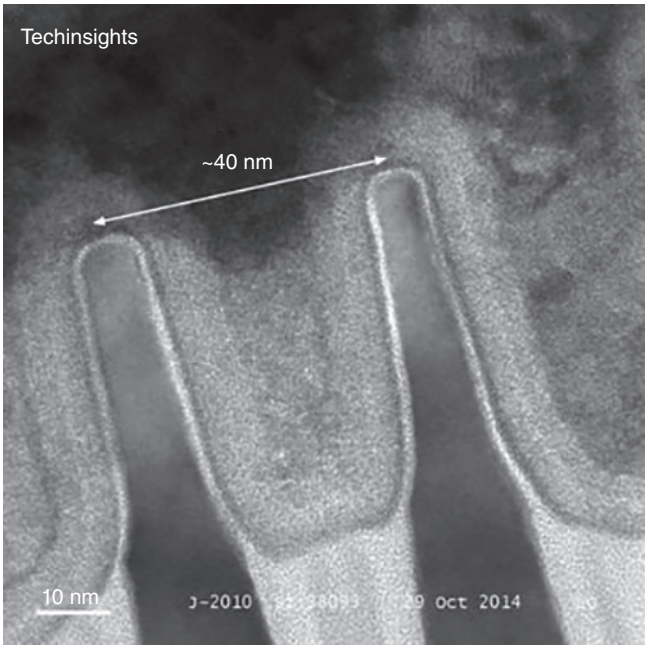

Figure 1.2 TEM image of Intel's 14-nm transistors with sub-40-nm fin pitch. Source: www.techinsights.com.

bulk MOSFETs is fully depleted silicon on insulator (FDSOI) MOSFETs, which reduce leakage between drain and source due to the removal of the substrate right below the channel [15]. The performance of the FDSOI MOSFETs is comparable with the double-gate field-effect transistors (DGFETs) in terms of SS, low junction capacitance, and high I_{on}/I_{off} ratio. Yet, the DGFETs have better scalability and can be manufactured on bulk Si wafers instead of silicon-on-insulator (SOI) wafers, which makes them more promising [16].

FinFETs or tri-gate FETs, which have three gates, have been found to be the most promising alternatives to MOSFETs due to their enhanced performance and simplicity of the fabrication process, which is compatible with and can be easily integrated into standard CMOS fabrication process (Figure 1.2) [17, 18]. In fact, an additional selective etch step is required in the FinFET fabrication process in order to create the third gate on top of the channel. FinFET devices have been explored carefully in the past decade. A large number of research articles that confirmed the improved short-channel behavior using different materials and processes have been published, as is shown in the following section. Next, the industrial state of FinFETs, their challenges, and limitations are discussed.

1.1.1 State-of-the-Art FinFETs

1.1.1.1 FinFET with Si Channel

In the semiconductor industry, silicon is the main channel material. The first FinFET technology (22-nm node) was produced by Intel in 2011. The second FinFET generation (14-nm node) published by Intel used strained Si channel [19].

The gate length was scaled from 26 to 20 nm in the second FinFET generation, which was possible due to new sub-fin doping and fin profile optimization. With a V_{DD} of 0.7 V, the saturation drive current is 1.04 mA μm^{-1} and the off current is 10 nA μm^{-1} for both nMOSFET (NMOS) and pMOSFET (PMOS). The SS is ~65 mV/decade, while the DIBL for N/PMOS is ~60/75 mV V^{-1}. High-density static random access memory (SRAM) having 0.0588 μm^2 cell size are also reported and fabricated using the 14-nm node. More recently, a research group from Samsung published a 7-nm CMOS FinFET using extreme ultraviolet (EUV) lithography instead of multiple-patterning lithography. This resulted in a reduction of the needed mask steps by more than 25%, in addition to providing smaller critical dimension variability and higher fidelity. The FinFET presented in this work consumes 45% less power and provides 20% faster speed than in the previous 10-nm technology. The reported SS is 65 and 70 mV/decade, and the DIBL is 30 and 45 mV V^{-1} for NMOS and PMOS, respectively. A 6T high-density and high-current SRAM memory has also been demonstrated using the 7-nm FinFET, and the results show a reduction in the bit line capacitance by 20% as a result of the reduction in the parasitic capacitance.

1.1.1.2 FinFET with High-Mobility Material Channel

The III–V materials gained growing attention for adoption as the channel material due to their promising characteristics such as high mobility, small effective mass, and, therefore, high injection velocity, in addition to near-ballistic performance. The first InGaSb pFET was demonstrated by Lu et al. [20], where a fin-dry etch technique was developed to obtain 15-nm narrow fins with vertical sidewalls. An equivalent oxide thickness (EOT) of 1.8 nm of Al_2O_3 was used as the gate oxide. The authors also demonstrated Si-compatible ohmic contacts that yielded an ultralow contact resistivity of 3.5×10^{-8} Ω cm^2. Devices with $L_g = 100$ nm and different fin widths (W_f) were demonstrated. The results show that with $W_f = 100$ nm, g_m of 122 μS μm^{-1} is achieved; while with $W_f = 30$ nm, g_m of 78 μS μm^{-1} is obtained.

Moreover, FinFETs with strained SiGe have lately attracted much interest due to their potential advantages such as higher mobility, built-in strain, and improved reliability with respect to conventional Si-based FETs. Very recently, a group of researchers at IBM demonstrated high-Ge-content strained SiGe FinFETs with replacement high-κ (HK)/metal gate (RMG). A long-channel subthreshold swing (SS) as low as ~68 mV/decade was reported [21]. This value is very competitive with other SiGe or Ge FinFETs with RMG process flow, where the reported SS values are in the range of 80–100 mV/decade [22]. In addition, a very high pFET hole mobility $\mu_{eff} = 235$ cm^2 V^{-1}s^{-1} was shown in a multi-fin device with average fin width of 4.6 nm and EOT of 7 Å which could be very promising for the sub-5-nm node FinFETs. Finally, in the same work, SiGe FinFETs with gate lengths $L_g = 25$ nm were fabricated using a gate-first flow. At a $V_{DD} = 0.5$ V, the devices showed DIBL = 40 mV, $SS_{lin}/SS_{sat} = 77/86$ mV/decade and $I_{on} = 430$ μA μm^{-1} at target high performance $I_{off} = 100$ nA μm^{-1}, which are among the largest reported values at such gate lengths (Figure 1.3).

In another work by Lei et al. [23], conducted in collaboration with Taiwan Semiconductor Manufacturing Co. (TSMC), the first GeSn FinFET device on

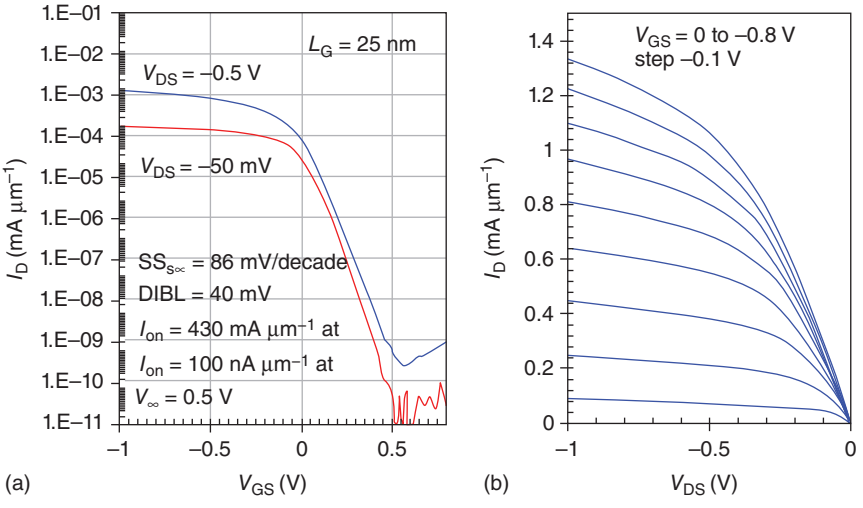

Figure 1.3 Transfer and output characteristics of high-Ge-content SiGe FinFETs with L_G 25 nm with gate first flow. Source: Hashemi et al. 2017 [21]. Reused with permision of IEEE.

a GeSnOI substrate was demonstrated with a channel length of 50 nm and $W_{Fin} = 20$ nm, and 4 nm HfO_2 was used as the gate oxide. The novel substrate was fabricated by the growth of high-quality GeSn by chemical vapor deposition (CVD) followed by a low-temperature process flow to get the GeSnOI. The GeSn pFET yielded the lowest SS of 79 mV/decade, the highest transconductance g_m of 807 µS µm^{-1}, and the highest hole mobility of 208 cm^2 V^{-1} s^{-1} (N_{inv} of 8×10^{12} cm^{-2}).

1.1.1.3 FinFET with TMD Channel

For sub-5-nm nodes, a body with sub-3-nm thickness is required to maintain good channel control. Most channel materials like the conventional Si or III–V face limitations in terms of mobility, quantum capacitance, or process at such ultrathin body (UTB) thickness. Advanced two-dimensional transition-metal dichalcogenide (TMD) is very promising in UTB thickness due to its sub-nanometer monolayer UTB thickness potential in addition to its good transport characteristics in nanometer thickness [24]. Chen et al. demonstrated the first 4-nm-thick TMD body FinFET with back gate control [25]. The main processes in the fabrication of the TMD FinFET is the compatibility of the CVD growth of TMD with CMOS processing, in addition to the reduction of the contact resistance by hydrogen plasma treatment of MoS_2. The V_t of this FinFET device can be adjusted dynamically by applying bias on the back gate. The front gate device showed an on/off current ratio over 10^5 with I_{on} of 200 µA µm^{-1} for $V_{dd} = 1$ V.

1.1.1.4 SOI versus Bulk FinFET

Bulk FinFETs are built on bulk-Si wafers, which are less expensive and have a lower defect density than do SOI wafers, while maintaining a better heat transfer rate to the Si substrate with respect to SOI FinFETs. The first Intel FinFET was a

bulk FinFET. Lee [1] studied the 14-nm node FinFET technology and compared bulk and SOI FinFET in terms of scalability, heat dissipation, and parasitic capacitance. Lee showed that both 14-nm FinFETs with bulk and SOI substrates have the same $I–V$ characteristics when the same geometry and doping concentration are used. Therefore, both devices have similar scalability. Moreover, the fins in bulk FinFETs are easily depleted, which allows for the reduction of the S/D to fin body junction capacitance to values that are lower than in the case of SOI FinFETs. Finally, to increase the heat transfer rate in SOI FinFETs, the buried oxide should be made thinner than 20 nm, which could have a negative impact on the device performance such as an increase in the parasitic capacitance.

Finally, it is worth mentioning that there are many other factors that affect the performance and reliability of FinFET devices, such as the materials used for metal gate/gate oxide, the shape of the fins (trapezoidal versus rectangular), the spacing between the fins, the fin edge roughness, choice of FET structure (lateral, vertical), and so on, which are not discussed in this chapter.

1.1.2 Industrial State

In 2011, Intel was the first company to use the 22-nm bulk FinFETs in mass production of central processing units (CPUs), which is 18% and 37% faster at 1 and 0.7 V, respectively, than Intel's 32-nm transistors [26]. Intel reported at the International Electron Devices Meeting (IEDM) that these 3D tri-gate transistors have a saturation current that exceeds 2 mA μm^{-1}. Several companies then followed Intel and announced the production of 3D transistors such as Samsung, TSMC, and Global Foundries. In 2015, Samsung announced the first production of the 14-nm FinFET-based transistors for mobile applications followed by the first mass production of the 10-nm FinFET (10LPE) in October 2016. Samsung was able to show improvements in power (40% lower power consumption than their 14-nm FinFETs), performance (27% higher performance), and scalability of the 3D tri-gate transistors (30% higher area efficiency).

However, at the 10-nm node, only three companies were capable of manufacturing such transistors: Samsung, Intel, and TSMC (Global Foundries excluded). Moreover, the geometries of the transistors produced at the leading manufacturers are different. For instance, the 10-nm FinFETs produced at TSMC and Samsung are denser than Intel's 14-nm FinFETs; however, they are closer to Intel's 14-nm FinFETs than they are to the Intel's 10-nm (the metal pitch in the Samsung's 10-nm is just 1 nm shorter than Intel's 14-nm).

In addition, some foundries use a hybrid node while others execute full node shrinking, which results in different geometries. In hybrid node shrinking, a new structure for the transistor (or a smaller transistor) is used (front end of line (FEOL)) but employing a set of design rules established previously for connecting transistors together (back end of line (BEOL)). In full node shrinking, both FEOL and BEOL are shrinking. In fact, TSMC and Samsung used the hybrid nodes at 16/14 nm where they introduced the new FinFET structure, while Intel is the only company executing full node shrinking with every new technology. It is worth noting that hybridized nodes allow the foundries to tackle a single set of challenges since the whole design process is not fully scaled down at once.

During Intel's Technology and Manufacturing Day 2017, Intel announced the mass production of its 10-nm process which used self-aligned quad patterning (SAQP) for the first time. Intel's 10-nm technology showed 45% less power consumption and 25% better performance than their 14-nm transistors with a minimum gate pitch of 54 nm (versus 70 nm for Intel's 14 nm) and a metal pitch of 36 nm (versus 52 nm for Intel's 14 nm). Also, Intel's 10-nm density is 2.7 × higher than the previous node (new density of 100.8 mega transistors mm^{-2}), with 25% taller (53-nm fin height) and more closely spaced fins (34-nm fin pitch).

Saumsung's 10 nm uses triple-patterning technology with a 68-nm contacted gate pitch, 51-nm metal pitch, dual-depth shallow trench isolation (STI) with a single dummy gate (ref Common Platform Alliance Paper which was presented in 2016), while TSMC's 10-nm used quad-patterning technology which allows a double increase in density compared to their 16-nm technology. TSMC claimed a poly pitch of 64 nm and a metal pitch of 42 nm with 35% less power consumption and 15% higher performance than their 16-nm technology.

In June 2017, Global Foundries announced the mass production of its 7-nm FinFET technology which offers 40% improvement in performance with volume production ramping in the second half of 2018. The initial production ramp of the 7-nm technology employs triple and quadruple patterning technology using a 193-nm excimer laser. Global Foundries will introduce EUV to its manufacturing process to accelerate the production ramp and improve the yield.

TSMC announced recently that its 7-nm FinFET will offer around 25% speed enhancement or a 35% power reduction over its 10-nm FinFETs, while Samsung announced the addition of the 8- and 6-nm process technologies to its current process roadmap with an aim of improving the cost competitiveness over its 10- and 7-nm technologies. It is also worth noting that Samsung's 7-nm will be its first technology to use EUV lithography.

1.1.3 Challenges and Limitations

The introduction of the FinFET technology has enabled the gate length scaling down to 7 nm with a 48-nm contacted poly pitch (CPP) due to improved device electrostatics [27]. The improved performance has been achieved through the "Fin Effect" boost (effective fin width/fin pitch) which increased the drive current for a certain capacitive load. However, the restrictions on the fin thickness are being rapidly approached, which would lead to a faster scaling in S/D sizes versus the contacted gate pitch. An increasing "Fin Effect" will thus result, which in combination with a plateau in the gate length would put pressure on the conduction path from contacts to S/D. In a work conducted by a group of researchers from Global Foundries and IBM, current contact resistivity of $\sim 2 \times 10^{-9}\ \Omega\,cm^2$ [28] will significantly deteriorate the performance of FinFETs below 40-nm CPP, while fully ohmic contacts with resistivity of $\sim 1 \times 10^{-10} \Omega\,cm^2$ [29] might push the CPP to below 30 nm. The work concluded that in order to further improve the performance and power consumption in future CMOS in the 30–40 nm CPP, industry will face pressure to use new device architectures or scaling choices [2].

Another challenge is that sub-5-nm nodes would need sub-3-nm body thickness for maintaining good channel control [25]. However, most of the channel

materials such as Si, Ge, and other III–V materials face fabrication, mobility, and quantum capacitance challenges at such small body thicknesses [30]. In addition, a group of researchers from IBM have fabricated test structures to unambiguously observe quantum confinement effects. The structures included fins with 40-nm fin pitch, 20-nm L_g, and 4- to 30-nm W_{Fin}. The measurements showed performance/mobility degradation, increase in series resistance, increase in variability, DIBL, and in V_t of NMOS/PMOS as the W_{Fin} is reduced [31], which confirms the challenges to be faced when further scaling down the FinFET technology.

Wavy FinFET has been proposed by Fahad et al. [32] as a promising structure for the high-performance technology node. The wavy transistor integrates 2D UTBs with the fin structure which maximizes the chip area utilization resulting in higher density, higher gain, and back bias capability. The structure was simulated using the 2013 International Technology Roadmap for Semiconductors (ITRS) specifications for the 7-nm node with UTB thickness of 2.5 nm and fin thickness of 6.8 nm. The authors reported an improved SS and DIBL performance of the wavy channel with 109% higher non-normalized ON-state drive performance as opposed to conventional FinFETs.

1.2 3D Integrated Circuit Technology

3D integration technology can denote either 3D packaging or 3D IC, which can be defined in different ways. In general, in 3D packaging, the vertical stacks are achieved via traditional methods of interconnects such as wire bonding and flip chip [33, 34]. However, in 3D IC, interconnections between different stacking layers are formed via through-silicon-vias (TSVs) [35]. Die stacking can be achieved by connecting separately manufactured dies or wafers vertically through one of three integration schemes: die-to-die, die-to-wafer, and wafer-to-wafer. The contacts (mechanical and electrical) can be achieved using either microbumps or by wire bonding as used in system-in-package (SIP) and package-on-package (POP) devices. Even though SIP is sometimes referred to as a 3D stacking technology, it is better referred to as a 2.5D technology. Another approach is to integrate dies horizontally on a silicon substrate using interposers. The benefits of using interposers are several: (i) lower communication power consumption due to the short communication distance between dies, (ii) the possibility of stacking separately manufactured dies from heterogeneous technologies to get the best out of all technologies, and (iii) enhanced yield and cost of the system due to the ability of fabricating and testing the smaller dies separately before integrating them into a silicon substrate instead of fabricating very large dies with much lower yield. The most promising approach of 3D integration is the monolithic approach, where active layers are vertically grown on top of each other and interconnects are made through TSVs which provide the densest connectivity.

There are several topics related to 3D integration that have recently gained a lot of attention in research. In the following, the main research topics with corresponding state-of-the-art technology are presented, followed by the industrial state and the main challenges of this technology.

1.2.1 Research State

1.2.1.1 Thermal Management

The biggest obstacle to the commercialization of 3D IC is the thermal management problem. As a matter of fact, the very thin thickness of chips in the 3D IC (<50 μm) in addition to the very high density of devices results in an increase in the temperature of the dies which are not close to the heat sink, and thereby deteriorating the performance of the system. In the past few years, research addressing thermal problems in 3D IC has gained growing attention. Goplen et al. reported that TSVs can act as a vertical path for heat flow [36]; therefore, thermal TSVs in addition to signal TSVs can be used to vertically transfer the heat and thereby reduce the die temperatures [37, 38]. Another study done by Lee et al. [39] showed that the heat transfer is directly proportional to the size of the via islands. In addition, it was found that a large number of TSVs can lead to routing congestion in the 3D ICs; thus, in addition to being expensive to fabricate, an optimization algorithm is needed to find the needed number of TSVs and their locations in order to be able to reduce the temperature of the dies. Moreover, Furumi et al. [40] proposed new cooling architectures for 3D ICs based on thermal sidewalls, interchip plates, and a bottom plate (thermal SIB). The experimental results conducted using a 3D thermal solver show that the thermal SIB can reduce the temperature in a 3D IC by over 40% when compared with structures that used a conventional heat sink only.

1.2.1.2 Through-silicon-vias

Using TSVs in 3D ICs and 3D packaging is very promising since it allows higher integration density, higher clock rate, and lower power dissipation [41]. In addition, TSVs are used in the 2.5D through-silicon interposers which enable the integration of heterogeneous dies on a silicon substrate. However, the fabrication of TSVs can be challenging: the etch process of the high-aspect-ratio TSVs should lead to scallop-free Si [26] and the Cu-filled TSV should be void-free [42]. This is in addition to challenges related to Cu protrusion affecting the BEOL reliability [43], thinning of TSV wafer [44], revealing of the backside of the TSV, and the bonding process [45]. In general, TSV fabrication requires the following steps: patterning of the via, etching the via, depositing the dielectric liner, metallization, and, finally, chemical-mechanical planarization (CMP) for planarization [46].

Currently, scaling down the TSVs is driven by the need to lower the thermal-mechanical stress in addition to its effect on the BEOL performance. The depth of the TSVs is limited, constrained by the wafer thinning (usually fixed at 50 μm). For a higher aspect ratio TSV (beyond 10 : 1), using the physical vapor deposition (PVD) barrier and seed process might lead to non-conformal films. IMEC and Lam Research Corp developed a low-cost process for getting conformal deposition of a very thin barrier and seed layer in high-aspect-ratio TSVs. The process consists of depositing a highly conformal thin oxide liner using atomic layer deposition (ALD), followed by the ALD deposition of the WN barrier, electroless plating NiB seed, and, finally, filling the TSV with copper using electrochemical deposition (ECD) [47]. Tokyo Electron Limited also reported another method to deposit highly conformal barrier and seed layers using electroless plating of Cu on

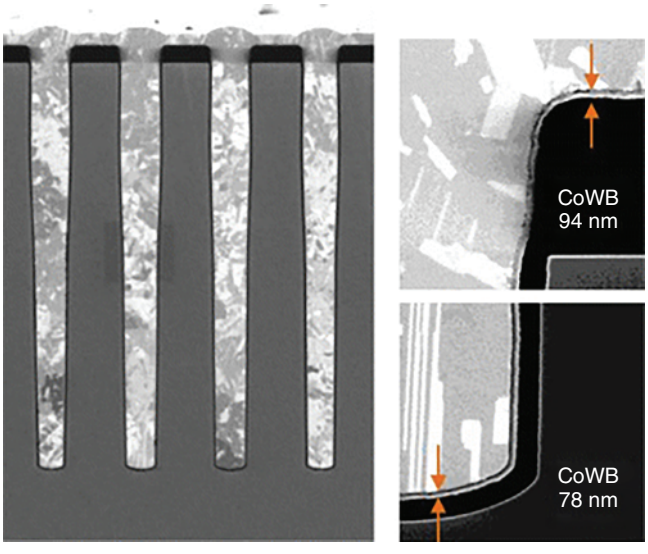

Figure 1.4 FIB-SEM after ECD-Cu filling on Eless-Cu/CoWB layers of 5 × 50 μm TSV. Source: Tanaka et al. 2015 [48]. Reused with permission of IEEE.

CoWB followed by Cu filling the TSV using ECD. Figure 1.4 shows the 5 × 50 μm TSV reported by Tokyo Electron Limited [48].

Another innovative metallization process was developed by Aveni (previously known as Alchimer). This metallization method is based on molecular engineering, where the film is grown molecule by molecule and can be applied in industry. First, a barrier layer is deposited by grafting and the NiB compound is used to make a Cu diffusion barrier which maintains the resistivity levels such that Cu can fill the high-aspect-ratio TSV using electrografting without the need for a cupper seed layer. The final fill process results in large, uniform, and high-purity grains of Cu, which could increase the yield due to eliminated voids, shorts, and opens [49].

1.2.1.3 Bonding in 3D IC

As already mentioned, the most important aspect of 3D IC is the ability to integrate heterogeneous dies fabricated at different foundries without performance degradation. The integration can be achieved either through wafer-on-wafer (WoW) bonding, chip-on-wafer (CoW) bonding, or chip-on-chip (CoC) bonding (Figure 1.5) [50]. WoW is the most preferred bonding due to its precise alignment [51]; more specifically, Cu metal-to-metal thermocompression bonding is the most favored among all bonding methods as it provides excellent electrical conductivity and mechanical strength after bonding [52]. During the thermocompression of Cu—Cu bonding, interdiffusion of Cu atom and grain growth across the bonding interface takes place. However, the main challenge to this process is to achieve it at low pressure and low temperature in order to avoid damaging the devices underneath or cause any reliability issues. But the Cu—Cu bonding process requires high temperature and pressure (or either of them) as native

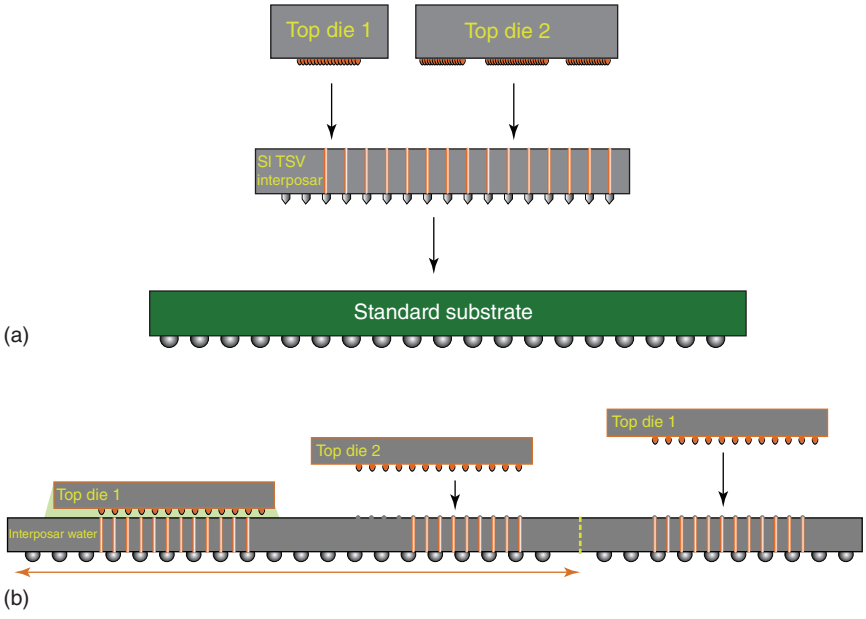

Figure 1.5 Assembly die stacking process flow: (a) CoS and (b) CoW. Source: https://amkor.com/.

oxide can be easily grown on the Cu surface, which inhibits the Cu interdiffusion and degrades the bonding quality [53]. Therefore, achieving high-quality Cu—Cu bonding at low temperature and pressure is needed in 3D IC.

Researchers have worked on several ways to avoid the surface oxidation of Cu and to remove the already grown native oxide. Shigetou et al. reported a bonding method based on surface-activated bonding (SAB). First, the native oxide is removed using argon bombardment at ultra-high vacuum (UHV), and then the bonding takes place at room temperature using a SAB flip-chip bonder [54]. However, the need for UHV increases the complexity of the process, and as a result becomes unattractive for manufacturing. Also, different chemistries have been proposed to do wet etching/cleaning of the native oxide such as hydrochloric acid [55], citric acid [56], sulfuric acid [57], and acetic acid [58]. Although some wet etching chemistries succeeded in removing the native oxide, immersing the wafer in such chemistries for a prolonged time might lead to etching the Cu and deteriorating the performance of the devices underneath. In another work conducted by Tan et al., a self-assembled monolayer (SAM) of alkane thiol, an organic monolayer, is deposited on the Cu surface to passivate it; and then the SAM is desorbed before the Cu—Cu bonding [59]. The use of the SAM passivation layer protected the Cu from growing native oxide, and the SAM removal process can be done at 250 °C [60]. However, all passivation based on SAM are not CMOS compatible.

In a work addressing this problem, a 3-nm Ti layer was used instead to passivate the Cu surface at 160 °C and 2.5 bar [61]. However, the Ti materials are challenging to be used in the damascene process; in addition, Ti can oxidize if exposed to air for more than two days, which is not favorable for the 3D IC process. In

a continuation to this work, it has been reported that a 3-nm of manganin alloy passivation layer deposited at 150 °C and low pressure led to a strong Cu—Cu bonding of 5 kN force, in addition to being damascene compatible.

1.2.1.4 Test and Yield

Every additional manufacturing step introduces a risk for defects and complicates the testing of the system. Yield is based on test results, and the cost is based on the test, yield, and throughput. 3D high yield is challenging to achieve, which is why the wire bonding of "known good dies" in 3D packages first found its application in mobile devices.

Any 3D IC process would be considered feasible only if its manufacturability yield is high. A group from Xilinx Inc. reported the key challenges faced during fabricating a 28-nm 3D IC with chip-on-wafer-on-substrate process [62]. During the initial ramp stage, most of the observed failures were related to the assembly at the interposer level such as open microbumps, opens and shorts in the interposer metal line, and TSV opens. Another failure mode is the deterioration of the transistors during the assembly of the 3D IC. The group developed a failure analysis technique based on a closed loop feedback, which resulted in improved yields.

1.2.2 Industrial State

Samsung is already using the monolithic approach to die stacking in 3D flash memory and smart sensors. The first commercial prototype of 3D IC (microcontroller) dates back to 2004 when Tezzaron released it [63]. In 2006, Intel assessed 3D chip stacking in Pentium 4 [64]. In 2011, IBM announced the introduction of the 3D chip production process [65]. Also, in 2012, Tezzaron released a prototype for its multicore design, which includes 64 core 3D-MAPS (*MA*ssively *P*arallel processor with *S*tacked memory) (http://arch.ece.gatech.edu/research/3dmaps/3dmaps.html)[66]. In 2013, a 128-Gb 3D NAND chip was introduced by Samsung which has 2× transistor density, 50% lower power consumption, 2× data storage speed, and 10× better retention characteristics compared to the planar version.

In 2015, Intel also introduced the 3D XPoint memory with 10× higher capacity than DRAM and 1000× faster than NAND flash [67]. Moreover, NVIDIA and AMD manufactured a high bandwidth memory (HBM) using 3D stacked memories, which is already used in the AMD GPU based on the Fiji architecture since 2015 (https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf). A high-performance RAM competing with HBM is the hybrid memory cube (HMC), which was introduced in 2011 by Micron and is based on DRAM stacked using TSVs (https://www.micron.com/products/hybrid-memory-cube). SanDisk and Toshiba announced in 2015 the production of the world's first 3D NAND with 48 layers and using BiCS (Bit-Cost Scalable) technology. The 3D NAND achieved 32 GB capacity with a storage of 3 bits per transistor. The latest version is called BiCS3, which will have 64 layers and will show a 40% larger capacity than the BiCS2, according to Toshiba.

Moreover, Micron reported the mass production of its 64-layer 3D NAND by the end of 2017, while Western Digital began mass production of its 64-layer

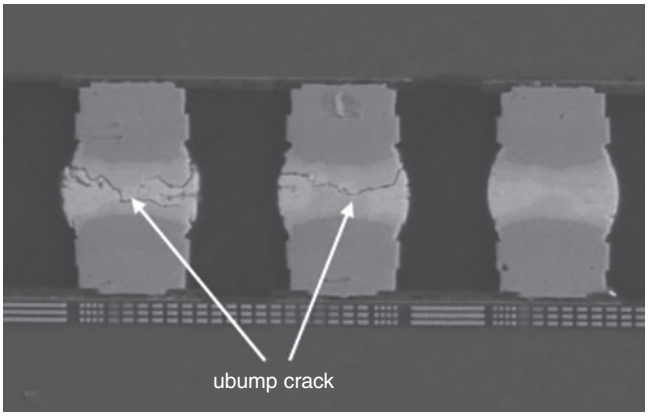

Figure 1.6 Picture of a microbump crack. Source: Yip et al. 2017 [68]. Reused with permission of IEEE.

3D NAND flash chips in 2017 (https://www.anandtech.com/show/10274/the-crucial-mx300-750gb-ssd-review-microns-3d-nand-arrives). Also, in early 2017, Intel announced the world's first commercial solid-state drive (SSD) based on 64-layer 3D NAND with a capacity of 512 GB (https://www.anandtech.com/show/11571/the-intel-ssd-545s-512gb-review-64layer-3d-tlc-nand-hits-retail).

1.2.3 Challenges and Limitations

Several challenges face the commercialization of 3D IC or 3D packaging. In a work done by Intel, it was found that yield estimates modeled using traditional methods can be pessimistic by as much as 50%. New analytical models have to be established to take into consideration other effects such as defect clustering and systematic defects introduced by equipment and handling issues during manufacturing. Moreover, it has been reported that the electrical performance of 3D IC with TSVs is affected due to the structure of the TSV with microbumps. TSV and microbump structures lead to local mechanical stress and strain due to the mismatch in the coefficient of thermal expansion (CTE) of Si, Cu TSV, and microbump (Figure 1.6) [68]. Moreover, crystal defects and stress can be induced in the Si chip as it is thinned down to less than a couple of tens of micrometers. Also, the gettering layers used to avoid the contamination of the metal and crystal defects can be removed from the Si chip as it is thinned down.

1.3 Neuromorphic Computing Technology

The flexibility of the VN architecture for "stored program" has led to enormous improvements in system performance for more than five years. However, since miniaturizing devices have slowed down in the past years, the energy and time used to transport data between memory and processor has become difficult, especially for data-centric applications such as real-time pattern recognition where state-of-the-art VN systems try hard to meet the performance of a human

being. The human brain outperforms advanced processors on many tasks such as unstructured data classification due to its parallel architecture connecting low-power neurons and synapses which act as computing and adaptive memory elements, respectively. The human brain performance is actually inspiring for novel non–VN computing models needed in future computing systems.

Even though designing neural circuits using electronic components dates back to the implementation of retinas [69] and perceptrons [70], modern research about very-large-scale integration (VLSI) technology using the nonlinear current characteristics began in the mid-1980s through collaboration between Richard Feynman, Carver Mead, Max Delbrück, and John Hopfield [71]. In fact, Mead tried to imitate the gradual synaptic transmission in the retina using the analog properties of transistors rather than operating them as digital switches. Mead was able to demonstrate that neuromorphic circuits using analog transistors instead of digital ones can match the physical properties of the proteic channels in neurons [72], leading to the need for a much smaller number of transistors to emulate neural systems.

In the neural system, neurons are connected to many other neurons, and they pass electrical and chemical signals to each other via synapses. These connections are either strengthened or weakened through a process called spike-timing-dependent plasticity (STDP), which is biologically observed [73, 74]. STDP changes depending on the timing between spikes (action potentials) within the input neuron (presynaptic) and output neuron (postsynaptic). In long-term potentiation (LTP), causal spiking strengthens synapses; while in long-term depression (LTD), the synaptic strength is weakened by causal spiking [75]. The change in the weight of synapses, also called synaptic plasticity, explains how the brain learns and memorizes [76].

Neuromorphic computing technology is considered a promising candidate for implementing applications such as self-learning, recognition of patterns, gestures, and speech using energy-efficient/low-power spiking networks. However, the progress in this technology faces two main challenges: (i) the lack of a full understanding of how the brain works and (ii) the lack of agreement on which technology can achieve synaptic and neural circuits with the best balance between cost, performance, and power consumption. Currently, a great deal of research is being conducted on different technologies for neuromorphic computing including mathematical and machine learning algorithms, neuromorphic datasets, field programmable gate array (FPGA) codes, photonic neuromorphic signal processing, nonvolatile memory (NVM) solutions, and so on. In this chapter, NVM for neuromorphic computing is discussed. In addition, the current industry state of neuromorphic computing, its challenges, and limitations are discussed.

1.3.1 State-of-the-Art Nonvolatile Memory as a Synapse

Around 10^{11} neurons and 10^{14} synapses exist in the human brain. In order to be able to implement brain-like processing architectures without using large and expensive areas on the silicon wafer, highly scalable and low-power memory devices are needed.

Different NVM devices have different physical properties and switching behaviors, and thus can be used to emulate synapses in different ways. For instance, when synapses are connected or not, an on/off NVM response would be sufficient; and this can be achieved using conductive-bridging random access memory (CBRAM). In other cases, synaptic weights are needed; therefore, an NVM with adjustable conductance would be required and this can be achieved using phase change memory (PCM) or memristor/resistive-random access memory (RRAM). In the following, the different types of NVM used to emulate synapses are briefly explained with state-of-the-art examples from the literature.

1.3.1.1 Phase Change Memory

In PCM, the state of the memory, whether programmed/SET or erased/RESET, depends on the difference in electrical resistivity between the amorphous and crystalline phases of the "phase change materials" leading to low (RESET) and high conductance (SET), respectively [77, 78] (Figure 1.7a).

PCM is attractive for neuromorphic applications where "device history" is needed, since the SET state can be achieved gradually by applying repetitive pulses to crystallize the phase of the plug in the device, resulting in a high-resistance state [84]. However, the RESET process can be only done sharply, since it involves melt and quench. The STDP can be implemented using a two-PCM approach: when an input neuron spikes, it outputs a signal (read pulse) and enters the LTP mode for a period of time t_{LTP}. If the postsynaptic neuron spikes during this period, a SET pulse is then sent to the LTP synapse. If not, then the LTD synapse is programmed, as shown in Figure 1.8a,b.

Suri et al. demonstrated that by adding a thin HfO_2 layer to the $Ge_2Sb_2T_5$ (GST)–based PCM, their synaptic performance can be improved [85, 86]. The addition of the interface layer affects the nucleation and growth activation energies, and thereby the crystallization kinetics, resulting in an increased dynamic range. In a later work, the authors developed a circuit model including

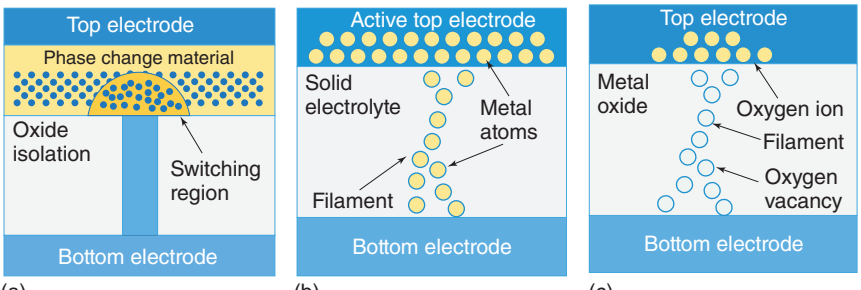

Figure 1.7 (a) Phase change memory (PCM) depends on the large difference in electrical resistivity between the amorphous (low-conductance) and crystalline (high-conductance) phases of so-called phase change materials [79, 80]. (b) Conductive-bridging RAM is based on the electrochemical formation of conductive metallic filaments through an insulating solid electrolyte or oxide [81]. (c) The conductive filaments in a filamentary RRAM are chains of defects through an otherwise insulating thin-film oxide [82]. Source: Reused with permission from [83].

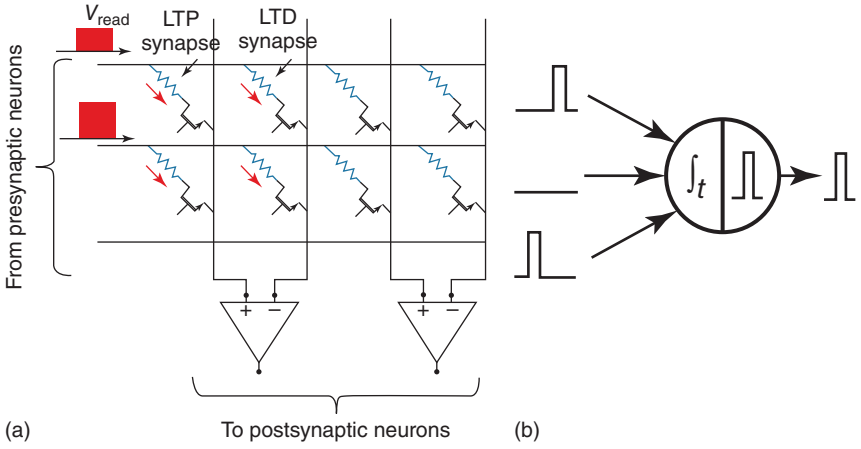

Figure 1.8 (a) Implementation of STDP with two-NVM-per-synapse scheme. Due to abrupt RESET in PCM devices, LTD and LTP are implemented with SET switching in different devices, with total weight of the synapse depending on the difference between these two conductances. (b) Key spiking characteristics of spiking neural network: downstream spikes depend on the time integration of continuous inputs, with synaptic weight change dependent on relative spike timing. Source: Reused with permission from [83].

the electrical and thermal characteristics of both top and bottom contacts with the PCM [79]. The authors showed that by enhancing the growth or nucleation rate, the maximum conductance can be reached in fewer pulses. They have also suggested that GST can offer more conductance states than GeTe, since GeTe (growth-dominated) saturated in conductance faster than the nucleation-dominated GST. Pattern learning and recognition was experimentally shown by Eryilmaz et al. using a 10×10 array of transistor-selected PCM cells [80]. They have also shown that longer training leads to lower initial resistance variation. Ambrogio et al. simulated larger networks of 28×28 pre- and 1 postneuron transistor-selected PCM cells (45-nm node) [87]. With two layers, the achieved MNIST (Modified National Institute of Standards and Technology) digit recognition probability was 33% with an error of 6%; while with three layers of networks, the recognition probability increased to 95.5% for 256 neurons with an error of 0.35%. The authors also demonstrated the capability of their network to learn new data in sequence and in parallel and forget previous data. Jackson et al. achieved STDP schemes using programming energies below 5 pJ in 10-nm pore PCM devices. Then, 100 leaky integrate-and-fire neurons were simulated. The authors showed the successful prediction of the next item in a sequence of four stimuli [88].

1.3.1.2 Conductive-Bridging RAM

Conductive-bridging random access memory (CBRAM) is based on the formation of conductive metallic filaments through an insulating solid electrolyte or oxide (Figure 1.7b) [77, 78]. CBRAM is promising for future NVM due to its characteristics such as extremely low power consumption (~nW), fast speed (~ns),

and scalability to the nanometer range. However, the SET state in the CBRAM is achieved abruptly as the formed filaments are quite conductive, leading to high currents for neuromorphic devices. Integrate-and-fire neurons would need larger capacitors.

STDP synaptic performance was achieved by Ohno et al. using an Ag_2S atomic switch [89]. The amplitude and widths of the pulses are found to affect the short-term memory formation. Then, the authors experimentally demonstrated the learning and forgetting mechanisms of two patterns using a 7×7 array of organic synapses [90]. Yu et al. demonstrated STDP with 1.5×10^8 cycles of depression and potentiation without noticeable degradation [91] using a 100×100 nm CBRAM-based memristor device that is connected to integrate and fire neurons. Using CBRAM devices as binary synapses and applying the STDP learning rule, Suri et al. demonstrated the recognition and extraction of real-time visual and audio patterns in an unsupervised manner [92]. Nonassociative and associative types of learning are shown in single $Pt/Ge_{0.3}Se_{0.7}/SiO_2/Cu$ memristive device by Ziegler et al. [93]. Sillin et al. developed a numerical model imitating the synapse-like properties of single atomic switches [94].

1.3.1.3 Filamentary RRAM

Filamentary resistive random access memory (F-RRAM) is similar to CBRAM; however, the conductive filament in this device is due to a chain of defects in an oxide once triggered by electrical field and/or local temperature increases rather than by metallic atoms (Figure 1.7c) [81]. F-RRAM is attractive because it requires metal-oxides, most of which are already used in CMOS fabrications, such as HfO_X, TiO_X, WO_X, TaO_X, FeO_X, and AlO_X in addition to laminates of such films. Adaptive synaptic changes leading to gradual memory have been demonstrated in such materials. The structure of an F-RRAM is based on a metal-insulator-metal structure which is CMOS compatible and highly scalable, in addition to achieving very low energy consumption per synaptic operation (sub-pJ), fast switching (<10 ns) [95], extremely small size (<10 nm), very low currents (1 μA programming current), and multibit storage [96]. However, similar to the CBRAM, the SET function is abrupt to the rapid formation of the filament.

A group of researchers at University of Michigan led by Prof. Wei Lu recently demonstrated a prototype memristor network to experimentally process natural images using the sparse-coding algorithm. In this study, a 16×32 sub-array from the 32×32 WO_x–based memristor array was used, corresponding to a 2× over-complete dictionary with 16 inputs and 32 output neurons and dictionary elements. The dictionary elements were learned offline using a realistic memristor model and an algorithm based on the "winner-take-all" (WTA) approach and Oja's learning rule. After training, they successfully experimentally reconstructed grayscale images using the 16×32 memristor crossbar [97]. Choi et al. demonstrated a multilevel RESET switching with continuously increasing RESET voltages in a GdO_x-based F-RRAM but with rapid SET switching [98]. Yu et al. and Wu et al. showed gradual switching in SET operation with constantly increasing external currents, and in RESET with uninterruptedly increasing reset voltages in $TiN/HfO_x/AlO_x/Pt$ and $TiN/Ti/AlO_x/TiN$ RRAM devices, respectively [99, 100].

Yu et al. used an F-RRAM with multilayer oxide-based Pt/HfO_x/TiO_x/HfO_x/TiO_x/TiN to achieve hundreds of resistance states during the RESET [101, 102]. Sub-pJ energy per spike was obtained with 10-ns short pulses. Finally, the multilevel resistance modulation was modeled using a stochastic model and was applied to a visual system simulation; a two-layer neural network was simulated using 1024 neurons and 16 348 oxide-based synapses, achieving up to 10% tolerance to resistance variations. Piccolboni et al. recently demonstrated an HfO_2-based vertical resistive random access memory (VRRAM) technology, each exhibiting two distinct states [103]. A stack of VRRAM devices forms a single synapse, with one common select transistor, exhibiting gradual conductance behavior. Simulation was used to demonstrate real-time auditory and visual pattern recognition.

1.3.2 Research Programs and Industrial State of Neuromorphic Computing

With the availability and advances in deep submicron CMOS technology, developing brain-like structures on electronic substrates has recently received growing attention, and large research projects on brain-like systems have been launched internationally. Currently, the two largest programs in this field worldwide are the SyNAPSE program (Systems of Neuromorphic Adaptive Plastic Scalable Electronics) in the United States (started in 2009, (http://www.artificialbrains.com/darpa-synapse-progra)) and the European Commission flagship Human Brain Project (started in 2013 (http://www.humanbrainproject.eu)). Funded by the Defense Advanced Research Projects Agency (DARPA), the SyNAPSE program aims to emulate a mammalian brain in terms of power consumption, size, and function using an electronic neuromorphic machine. Then, robots with the intelligence of cats and mice would be built using such artificial brains. The neuromorphic microprocessor should be able to simulate the activity of 10 billion neurons and 100 trillion synapses using less than two liters of space and 1 kW of power (http://www.artificialbrains.com/darpa-synapse-progra). A project funded by DARPA's SyNAPSE initiative is the "Cognitive Computing via Synaptronics and Supercomputing" (C2S2) program, which is headed by IBM. A remarkable outcome of this project is the "True North chip," which is the largest chip fabricated at IBM and the second largest CMOS chip worldwide. This chip includes a 64×64 network of cores for digital applications, 256 millions of programmable synapses, and over 400 million bits of on-chip SRAM memory as storage space for neuron and synapse parameters. The 28-nm CMOS technology node with a die size of 4.3 cm^2 is used to fabricate the 5.4 billion transistors on the chip. The "True North chip" consumes 70 mW power (or 20 mW cm^{-2}), which is comparable to the cortex; however, the conventional CPU consumes at least 3 orders of magnitude higher power (50–100 W cm^{-2}) [104].

The Human Brain Project (HBP) is a European Commission (EC) flagship project with goals of increasing world awareness about the fields of neuroscience and brain-related medicine. This program has several subprojects, and one of them (called SP9) aims to develop a neuromorphic computing system using (i) physical brain-emulation models (with 200 000 neurons fabricated using 180-nm

CMOS technology), (ii) real-time numerical models (with 18-Advanced Reduced instruction set computer Machines (ARM) cores fabricated using the 130-nm CMOS technology, and (iii) software tools to design, run, and record the performance of the system [104]. The Blue Brain Project (launched in 2005 and led by EPFL and IBM) aims to understand the structure and functionality of the brain using simulations of the rodent and the brain. The simulations are conducted using an IBM supercomputer (Blue Gene, 10TB) with 8K CPUs to simulate artificial neural networks (http://bluebrain.epfl.ch/page-56882-en.html). Closely related to this project is the BrainScaleS (brain-inspired multiscale computation in neuromorphic hybrid systems), which is European funded. The BrainScaleS project uses Petaflop supercomputers to run numerical simulations to emulate and understand the brain-information processing. The hardware consists of the HICANN (High Input Count Analog Neural Network) chip, which has 112K synapses and 512 neuron circuits fabricated in a 180-nm CMOS technology (http://brainscales.kip.uni-heidelberg.de). Another impressive neuromorphic computing project is the SpiNNaker project [96], which consists of multiple core chips with multi-ARM interconnected through a specific communication technology. An 18-ARM9 CPU is included in each SpiNNaker package with a DRAM memory of 128 Mbyte, and each ARM core can real- time simulate 1000 neurons. The current full SpiNNaker board consists of 47 packages with goals of assembling 1200 boards with 90-kW power consumption.

1.4 Quantum Computing Technology

Yuri Manin and Richard Feynman independently reported that simulating physics using quantum computers would be more beneficial than using classical computers. Other than simulating physics, the question arose regarding whether quantum computers could outperform classical computers in solving other problems too. Paul Benioff and David Deutsch [105] later designed a layout for the quantum computer, while P. Shor and L. Grover developed the first algorithms that could run more efficiently on such quantum computers than on classical ones [106, 107].

In classical computers, the unit of information is the bit, which exists in two states: 0 and 1. The computations in such computers are a sequence of operations known as gates which are applied to bits. The computer's size and clock rate vary with the physical medium in which the bits are stored; however, the computational power of the computer is not affected by the bits' physical medium. Thus, two computers with the same storage capacity (bits) and set of operations (gates) are considered equivalent. In quantum computing, however, the unit of information is called "qubit" and the relevant operations are the "quantum gates." Unlike the bit, the qubit can exist in the state |0>, |1> (labeled using Dirac's "bra-ket") or a superposition of the two states. Different approaches are used to design the physical medium of the qubit. Nevertheless, approaches based on semiconductors are gaining growing attention since they can be produced easily using lithography technology. The favored quantum degree of freedom in semiconductors is the spin since it does not interact with the environment. To be specific,

silicon is an excellent candidate for spin qubits since it can be chemically purified, resulting in long-spin coherence time (in the seconds range) [108–110].

Different schemes have been proposed to implement qubits and quantum gates such as optics, ion traps, and nuclear spins in nuclear magnetic resonance devices. All of these schemes face several challenges and are still under development. Other researchers are focusing on developing advanced algorithms and mathematical models to run quantum computers. In this chapter, the qubits based on spins and superconducting materials are discussed.

1.4.1 Quantum Bit Requirement

The quantum bit implementation requires a system that can hold two states 0 and 1, and that can be initialized, acted on, and read [111]. Unlike the conventional electronics where the bits are transferred through wires from the processor to the memory, the qubits actually do not move; however, the control signals (logical gates) are brought close to the qubits to operate on and control them. Like digital electronics, an arbitrary logic can be implemented using a discrete set of logical operations [112]. At least two qubits are needed to be acted on at the same time by the set of operations, and the state of one qubit affects the state of the other. Therefore, computation requires qubits that can be coupled in a scalable manner and with high fidelity. Solid-state approaches are promising for the integration of a large number.

1.4.2 Research State

Quantum computers are able to solve problems related to chemistry, materials science, and mathematics that are beyond the capabilities of any supercomputer. The power of the quantum computers arises from the nature of the quantum bits that can exist in both states 0 and 1 at the same time, which is called the quantum superposition state. As a result, the computing power doubles with each additional qubit. Promising areas of research include superconducting circuits, electron spins in impurities, electron spins in semiconductor quantum dots, single photons [113], trapped ions [114], single defects or atoms in diamond [115, 116] and silicon [117], and so on, with single-qubit fidelities exceeding the threshold needed for fault-tolerant quantum computing.

Here, the two most promising systems which are the most similar to current solid-state circuits are discussed: superconducting circuits and electron spins in semiconductor quantum dots. It is worth mentioning that both of these qubits require cryogenic temperatures for operation, depend on analog control signals, and use radio frequency (RF) circuits to read the qubit state.

1.4.2.1 Spin-Based Qubits

Spin qubits are based on the intrinsic properties of semiconductors, such as electron spins trapped in the potential of chemical impurity or quantum dot. Spins are indeed protected from charge noise as a result of the weak spin-orbit coupling.

Loss and DiVincenzo focused on semiconductor quantum dots patterned using lithography. They reported the initialization of the ground state of the spin at low temperatures and high magnetic fields, the control of the spin using the electron spin resonance (ESR) toolbox, and the read based on the spin-to-charge conversion process [118]. The electrical control of the overlap in wave function results in an exchange coupling that can be tuned by the gate voltage. When combined with ESR, a controlled-not (CNOT) gate can be implemented, which is an essential logic operation in the implementation of a quantum computer. This was first demonstrated in GaAs quantum dots [119, 120]. Also, a high-fidelity two-qubit gate was recently demonstrated in a silicon device [121].

In addition, latest experiments have reported that the lifetime of the electron spin limits the high-fidelity readout of the qubits. Using a nanodevice, T. Watson et al. reported the longest lifetime of any electron spin qubit (30 seconds). The researchers engineered the electron wave function within phosphorous atom quantum dots such that the spin relaxation is minimized. Due to the longer lifetimes of the electron spin, the authors reported the readout of two sequential qubits with 99.8% fidelities, which are above the surface-code fault-tolerant threshold [122].

In another work, Veldhorst et al. reported the control over the spin states of the qubits by applying voltages with GHz frequencies. The authors used a phosphorous single-atom transistor with all epitaxial monolayer-doped gates (Figure 1.9a,b) and pulsed spectroscopy with selective transport via excited states which enabled the differentiation between the excited states of the single P atom. [121]

1.4.3 Superconducting Circuits for Quantum Information

Superconducting quantum circuits consist of a high number of atoms (usually aluminum) assembled with metallic wire/plate shapes and are based on the electric LC oscillator [121]. Two phenomena form the basis for the operation of the superconducting qubits: (i) superconductivity, which is the frictionless flow of electrical fluid through metals at low temperatures, and (ii) the Josephson effect, which provides nonlinearity to the circuit without causing dephasing or dissipation. The electron fluid motion around the circuit is denoted with the flux F reaching the inductor, which acts as the center-of-mass position in a mass-spring mechanical oscillator [123]. The Josephson tunnel junction converts the circuit into an artificial atom which can be selectively excited from the ground state to an excited state and used as a qubit. By changing the relative strengths of the three energies associated with the capacitance, inductance and tunnel element, different shapes of potential energies can be achieved. The performance of the qubits has drastically enhanced as the fabrication, measurements, and materials affecting coherence have been understood and enhanced. Moreover, other design variations have been introduced such as quantronium [124], fluxonium [125], and hybrid qubits [126], all of which are fabricated using the same materials but aim to enhance the performance by lowering their sensitivity to decoherence mechanisms in the environment.

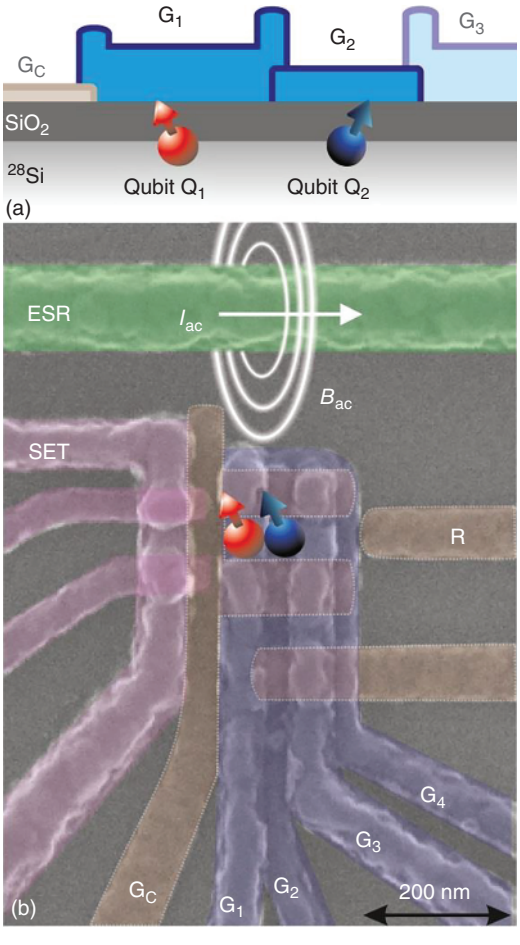

Figure 1.9 Silicon two-qubit logic device, incorporating SET readout and selective qubit control. Schematic (a) and scanning electron microscope colored image (b) of the device. The quantum dot structure (labels G_C and G_{1-4}) can be operated as a single or double quantum dot by appropriate biasing of gate electrodes G_1–G_4, where the dots $D_{1,2}$ are confined underneath gates $G_{1,2}$, respectively. The confinement gate G_C runs underneath G_1–G_3 and confines the quantum dot on all sides except on the reservoir (R) side. Qubit operation is achieved via an ac current I_{ac} through the ESR line, resulting in an ac magnetic field B_{ac}. Source: Devoret and Schoelkopf 2013 [123]. Reused with permission of Nature Publishing Group.

1.4.4 Industry State

In March 2017, IBM introduced two of its most powerful quantum computing processors for the IBM Q to help researchers and scientists solve problems that are not possible with today's most powerful computer. The two new IBM quantum processors include the following:

- A processor with 16 qubits which will enable solving of more complex experimentations than the previous 5-qubit processor.
- A processor with 17 qubits which is the first commercial prototype from IBM. This processor is the most powerful quantum processor invented by IBM to date: it is at least 2× more powerful than what is available to users on the IBM Cloud (https://phys.org/news/2017-05-ibm-powerful-universal-quantum-processors.html).

Also, in October 2017, Intel announced its new 17-qubit chip, which was delivered to QuTech in the Netherlands. It is worth mentioning that these quantum

computers still cannot compete with current classical computers; however, the future is bright, especially with superconducting qubits (https://newsroom.intel.com/press-kits/quantum-computing/).

Moreover, IBM and Intel are not the only two companies working on building quantum computers. Google is also preparing a 50-qubit quantum computer, which is going to be used to solve a scientific previously unsolvable problem. Other companies working on creating quantum computers include Tigetti Computing, which is a startup in Berkeley, CA, and Microsoft Corp ((https://www.sciencealert.com/google-s-quantum-announcement-overshadowed-by-something-even-bigger) and (https://news.microsoft.com/features/new-microsoft-breakthroughs-general-purpose-quantum-computing-moves-closer-reality/)).

In addition, the European Commission is funding a €1 billion flagship project on quantum computing to launch in 2018 (https://ec.europa.eu/digital-single-market/en/news/european-commission-will-launch-eu1-billion-quantum-technologies-flagship).

1.4.5 Challenges and Limitations to Quantum Computing

Some of the challenges facing quantum computing technology are discussed in this section. First of all, there is the need for quantum error correction since the qubit states change in time in uncontrolled ways due to their interaction with the environment (aka decoherence). The error probability calculated by the quantum error correction algorithm must be below 1% (accuracy threshold for fault tolerance) [123]. This would require additional qubits for encoding and decoding. However, the number of qubits needed should be reduced; in fact, an estimated number of qubits needed to compute a molecule reaches millions. Therefore, this number should be brought down by several orders of magnitude.

Moreover, specific electronics should be built to produce the control signals, and to store and process the output signals. These electronics include analog-to-digital converters (ADCs), digital-to-analog converters (DACs), RF sources, amplifiers, multiplexer circuits, digital data processing units, and so on. The electronics need to be low cost (at $1.00 per qubit) and show high accuracy (exceeding the 1% accuracy threshold by 2 orders of magnitude). In addition, some electronics might require cryogenic temperatures to operate, which poses a tight power budget. Also, qubits receive control signals from outside; therefore, multiplexing strategies must be employed in the interconnect technology between the qubits and the control and output electronics [127].

References

1 Lee, J.H. (2016). Bulk FinFETs: design at 14 nm node and key characteristics. In: *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, 33–64. Netherlands: Springer.

2 Razavieh, A., Zeitzoff, P., Brown, D.E. et al. (2017). Scaling challenges of FinFET architecture below 40nm contacted gate pitch. In: 2017 75th Annual Device Research Conference (DRC), pp. 1–2.

3 Joyner, J.W., Venkatesan, R., Zarkesh-Ha, P. et al. (2001). Impact of three-dimensional architectures on interconnects in gigascale integration. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 9 (6): 922–928.
4 Beyne, E. (2006). The rise of the 3rd dimension for system intergration. In: 2006 International IEEE Interconnect Technology Conference, pp. 1–5.
5 Pershin, Y.V. and Di Ventra, M. (2011). Solving mazes with memristors: a massively parallel approach. *Physical Review E* 84 (4): doi: 10.1103/physreve.84.046703.
6 Pickett, M.D., Medeiros-Ribeiro, G., and Williams, R.S. (2013). A scalable neuristor built with Mott memristors. *Nature materials* 12 (2): 114–117.
7 Pershin, Y.V. and Di Ventra, M. (2012). Neuromorphic, digital, and quantum computation with memory circuit elements. *Proceedings of the IEEE* 100 (6): 2071–2080.
8 Hu, C. (1996). Gate oxide scaling limits and projection. In: International Electron Devices Meeting, 1996, IEDM'96, pp. 319–322. IEEE.
9 Yeo, Y.C., King, T.J., and Hu, C. (2003). MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations. *IEEE Transactions on Electron Devices* 50 (4): 1027–1035.
10 Chen, J., Chan, T.Y., Chen, I.C. et al. (1987). Subbreakdown drain leakage current in MOSFET. *IEEE Electron Device Letters* 8 (11): 515–517.
11 Ferain, I., Colinge, C.A., and Colinge, J.P. (2011). Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature* 479 (7373): 310–316.
12 Wong, H.S., Chan, K.K., and Taur, Y. (1997). Self-aligned (top and bottom) double-gate MOSFET with a 25 nm thick silicon channel. In: International Electron Devices Meeting, 1997, IEDM'97, Technical Digest, pp. 427–430. IEEE.
13 Choi, Y.K., Lindert, N., Xuan, P. et al. (2001). Sub-20 nm CMOS FinFET technologies. In: International Electron Devices Meeting, 2001, IEDM'01. Technical Digest, pp. 421–424. IEEE.
14 Mitard, J., Witters, L., Loo, R. et al. (2014). 15nm-W FIN high-performance low-defectivity strained-germanium pFinFETs with low temperature STI-last process. In: 2014 Symposium on VLSI Technology (VLSI-Technology), Digest of Technical Papers, pp. 1–2. IEEE.
15 Choi, Y.K., Asano, K., Lindert, N. et al. (1999). Ultra-thin body SOI MOSFET for deep-sub-tenth micron era. In: International Electron Devices Meeting, 1999, IEDM'99, Technical Digest, pp. 919–921. IEEE.
16 Doris, B., Cheng, K., Khakifirooz, A. et al. (2013). Device design considerations for next generation CMOS technology: Planar FDSOI and FinFET. In: 2013 International Symposium on VLSI Technology, Systems, and Applications (VLSI-TSA), pp. 1–2. IEEE.
17 Auth, C. (2012). 22-nm fully-depleted tri-gate CMOS transistors. In: 2012 IEEE Custom Integrated Circuits Conference (CICC), pp. 1–6. IEEE.
18 Guillorn, M., Chang, J., Bryant, A. et al. (2008). FinFET performance advantage at 22nm: An AC perspective. In: 2008 Symposium on VLSI Technology, pp. 12–13. IEEE.

19 Natarajan, S., Agostinelli, M., Akbar, S. et al. (2014). A 14nm logic technology featuring 2 nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm 2 SRAM cell size. In: 2014 IEEE International Electron Devices Meeting (IEDM), pp. 3–7. IEEE.

20 Lu, W., Kim, J.K., Klem, J.F. et al. (2015). An InGaSb p-channel FinFET. In: 2015 IEEE International Electron Devices Meeting (IEDM), pp. 31–36. IEEE.

21 Hashemi, P., Ando, T., Balakrishnan, K. et al. (2017). High performance PMOS with strained high-Ge-content SiGe fins for advanced logic applications. In: 2017 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), pp. 1–2. IEEE.

22 Hashemi, P., Ando, T., Balakrishnan, K. et al. (2016). Replacement high-K/metal-gate High-Ge-content strained SiGe FinFETs with high hole mobility and excellent SS and reliability at aggressive EOT ~ 7Å and scaled dimensions down to sub-4nm fin widths. In: 2016 IEEE Symposium on VLSI Technology, pp. 1–2. IEEE.

23 Lei, D., Lee, K.H., Bao, S. et al. (2017). The first GeSn FinFET on a novel GeSnOI substrate achieving lowest S of 79 mV/decade and record high Gm, int of 807 μS/μm for GeSn P-FETs. In: 2017 Symposium on VLSI Technology, pp. T198–T199. IEEE.

24 Lee, Y.J., Luo, G.L., Hou, F.J. et al. (2016). Ge GAA FETs and TMD FinFETs for the applications beyond Si: a review. *IEEE Journal of the Electron Devices Society* 4 (5): 286–293.

25 Chen, M.C., Li, K.S., Li, L.J. et al. (2015). TMD FinFET with 4 nm thin body and back gate control for future low power technology. In: 2015 IEEE International Electron Devices Meeting (IEDM), pp. 32–2. IEEE.

26 Morikawa, Y., Murayama, T., Sakuishi, Y.N.T. et al. (2013). Total cost effective scallop free Si etching for 2.5 D & 3D TSV fabrication technologies in 300mm wafer. In: 2013 IEEE 63rd Electronic Components and Technology Conference (ECTC), pp. 605–607. IEEE.

27 Xie, R., Montanini, P., Akarvardar, K. et al. (2016). A 7nm FinFET technology featuring EUV patterning and dual strained high mobility channels. In: 2016 IEEE International Electron Devices Meeting (IEDM), pp. 2–7. IEEE.

28 Niimi, H., Liu, Z., Gluschenkov, O. et al. (2016). Sub- 10^{-9} Ω -cm^2 n-type contact resistivity for FinFET technology. *IEEE Electron Device Letters* 37 (11): 1371–1374.

29 Maassen, J., Jeong, C., Baraskar, A. et al. (2013). Full band calculations of the intrinsic lower limit of contact resistivity. *Applied Physics Letters* 102 (11): 111605.

30 Liu, W., Kang, J., Cao, W. et al. (2013). High-performance few-layer-MoS_2 field-effect-transistor with record low contact-resistance. In: 2013 IEEE International Electron Devices Meeting (IEDM), pp. 19–4. IEEE.

31 Chang, J.B., Guillorn, M., Solomon, P.M. et al. (2011). Scaling of SOI FinFETs down to fin width of 4 nm for the 10nm technology node. In: 2011 Symposium on VLSI Technology (VLSIT), pp. 12–13. IEEE.

32 Fahad, H.M., Hu, C., and Hussain, M.M. (2015). Simulation study of a 3-D device integrating FinFET and UTBFET. *IEEE Transactions on Electron Devices* 62 (1): 83–87.

33 Mahajan, R., Sankman, R., Patel, N. et al. (2016). Embedded multi-die interconnect bridge (EMIB)--a high density, high bandwidth packaging interconnect. In: 2016 IEEE 66th Electronic Components and Technology Conference (ECTC), pp. 557–565. IEEE.

34 Zhang, D. and Lu, J.J.Q. (2017). 3D integration technologies: an overview. In: *Materials for Advanced Packaging*, 1–26. Springer International Publishing.

35 Patti, R.S. (2006). Three-dimensional integrated circuits and the future of system-on-chip designs. *Proceedings of the IEEE* 94 (6): 1214–1224.

36 Goplen, B. and Sapatnekar, S. (2005). Thermal via placement in 3D ICs. In: Proceedings of the 2005 international symposium on Physical design, pp. 167–174. ACM.

37 Kandlikar, S.G. and Ganguly, A. (2017). Fundamentals of heat dissipation in 3D IC packaging. In: *3D Microelectronic Packaging*, 245–260. Springer International Publishing.

38 Cong, J., Wei, J., and Zhang, Y. (2004). A thermal-driven floorplanning algorithm for 3D ICs. In: IEEE/ACM International Conference on Computer Aided Design, 2004. ICCAD-2004, pp. 306–313. IEEE.

39 Lee, S., Lemczyk, T.F., and Yovanovich, M.M. (1992). Analysis of thermal vias in high density interconnect technology. In: Eighth Annual IEEE Semiconductor Thermal Measurement and Management Symposium, 1992. SEMI-THERM VIII., pp. 55–61. IEEE.

40 Furumi, K., Imai, M., and Kurokawa, A. (2017). Cooling architectures using thermal sidewalls, interchip plates, and bottom plate for 3D ICs. In: 2017 18th International Symposium on Quality Electronic Design (ISQED), pp. 283–288. IEEE.

41 Karnezos, M., Carson, F., Pendse, R., and ChipPAC, S.T.A.T.S. (2005). 3D packaging promises performance, reliability gains with small footprints and lower profiles. *Chip Scale Review* 1: 29.

42 Wolf, M.J., Dretschkow, T., Wunderle, B. et al. (2008). High aspect ratio TSV copper filling with different seed layers. In: 58th Electronic Components and Technology Conference, 2008. ECTC 2008. pp. 563–570. IEEE.

43 Che, F.X., Putra, W.N., Heryanto, A. et al. (2013). Study on Cu protrusion of through-silicon via. *IEEE Transactions on Components, Packaging and Manufacturing Technology* 3 (5): 732–739.

44 Che, F.X. (2014). Dynamic stress modeling on wafer thinning process and reliability analysis for TSV wafer. *IEEE Transactions on Components, Packaging and Manufacturing Technology* 4 (9): 1432–1440.

45 Huang, B.K., Lin, C.M., Huang, S.J. et al. (2013). Integration challenges of TSV backside via reveal process. In: 2013 IEEE 63rd Electronic Components and Technology Conference (ECTC), pp. 915–917. IEEE.

46 Redolfi, A., Velenis, D., Thangaraju, S. et al. (2011). Implementation of an industry compliant, $5 \times 50\mu m$, via-middle TSV technology on 300mm wafers. In: 2011 IEEE 61st Electronic Components and Technology Conference (ECTC), pp. 1384–1388. IEEE.

47 Van Huylenbroeck, S., Li, Y., Heylen, N. et al. (2015). Advanced metallization scheme for $3 \times 50\mu m$ via middle TSV and beyond. In: 2015 IEEE 65th

Electronic Components and Technology Conference (ECTC), pp. 66–72. IEEE.

48 Tanaka, T., Iwashita, M., Toshima, T. et al. (2015). Electro-less barrier/seed formation in high aspect ratio via. In: 2015 IEEE 65th Electronic Components and Technology Conference (ECTC), pp. 78–82. IEEE.

49 3D TSVs, aveni (2016). http://aveni.com/wet-deposition/3d-tsvs/ (accessed 31 May 2018).

50 Lee, S.H., Chen, K.N., and Lu, J.J.Q. (2011). Wafer-to-wafer alignment for three-dimensional integration: a review. *Journal of Microelectromechanical Systems* 20 (4): 885–898.

51 Lu, J.Q., McMahon, J.J., and Gutmann, R.J. (2012). Hybrid metal/polymer wafer bonding platform. In: *Handbook of Wafer Bonding*, 215–236. Wiley-VCH.

52 Cho, S. (2011). Technical challenges in TSV integration to Si. In: Sematech Symposium Korea, pp. 1–33.

53 Pangracious, V., Marrakchi, Z., and Mehrez, H. (2015). Three-dimensional integration: a more than moore technology. In: *Three-Dimensional Design Methodologies for Tree-based FPGA Architecture*, 13–41. Springer International Publishing.

54 Shigetou, A., Itoh, T., and Suga, T. (2006). Bumpless interconnect of Cu electrodes in millions-pins level. In: 56th Electronic Components and Technology Conference, 2006. Proceedings. pp. 4. IEEE.

55 Chen, K.N., Tan, C.S., Fan, A., and Reif, R. (2005). Copper bonded layers analysis and effects of copper surface conditions on bonding quality for three-dimensional integration. *Journal of Electronic Materials* 34 (12): 1464–1467.

56 Swinnen, B., Ruythooren, W., De Moor, P. et al. (2006). 3D integration by Cu-Cu thermo-compression bonding of extremely thinned bulk-Si die containing 10 μm pitch through-Si vias. In: International Electron Devices Meeting, 2006. IEDM'06. pp. 1–4. IEEE.

57 Huffman, A., Lannon, J., Lueck, M. et al. (2009). Fabrication and characterization of metal-to-metal interconnect structures for 3-D integration. *Journal of Instrumentation* 4 (03): P03006.

58 Fan, J., Lim, D.F., and Tan, C.S. (2013). Effects of surface treatment on the bonding quality of wafer-level Cu-to-Cu thermo-compression bonding for 3D integration. *Journal of Micromechanics and Microengineering* 23 (4): 045025.

59 Tan, C.S., Lim, D.F., Singh, S.G. et al. (2009). Cu–Cu diffusion bonding enhancement at low temperature by surface passivation using self-assembled monolayer of alkane-thiol. *Applied Physics Letters* 95 (19): 192108.

60 Lim, D.F., Wei, J., Leong, K.C., and Tan, C.S. (2013). Cu passivation for enhanced low temperature ($\leqslant 300\,^{\circ}C$) bonding in 3D integration. *Microelectronic Engineering* 106: 144–148.

61 Panigrahi, A.K., Bonam, S., Ghosh, T. et al. (2016). Ultra-thin Ti passivation mediated breakthrough in high quality Cu-Cu bonding at low temperature and pressure. *Materials Letters* 169: 269–272.

62 Chaware, R., Hariharan, G., Lin, J. et al. (2015). Assembly challenges in developing 3D IC package with ultra high yield and high reliability. In: 2015 IEEE 65th Electronic Components and Technology Conference (ECTC), pp. 1447–1451. IEEE.

63 Tezzaron 3D-IC Microcontroller Prototype [Online]. (2016). http://www.tachyonsemi.com/OtherICs/3D-IC_8051_prototype.htm (accessed 11 February 2016).

64 Black, B., Annavaram, M., Brekelbaum, N. et al. (2006, December). Die stacking (3D) microarchitecture. In: *Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture*, 469–479. IEEE Computer Society.

65 IBM Press Release [Online], in German. http://www-03.ibm.com/press/de/de/pressrelease/36129.wss (accessed 11 February 2016).

66 Kim, D.H., Athikulwongse, K., Healy, M. et al. (2012). 3D-MAPS: 3D massively parallel processor with stacked memory. In: 2012 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), pp. 188–190. IEEE.

67 Intel® Optane™ (2016). Supersonic memory revolution to take-off in 2016. http://www.intel.eu/content/www/eu/en/it-managers/non-volatile-memory-idf.html (accessed 11 February 2016).

68 Yip, L., Hariharan, G., Chaware, R. et al. (2017). Board level reliability optimization for 3D IC packages with extra large interposer. In: 2017 IEEE 67th Electronic Components and Technology Conference (ECTC), pp. 1269–1275. IEEE.

69 Fukushima, K., Yamaguchi, Y., Yasuda, M., and Nagata, S. (1970). An electronic model of the retina. *Proceedings of the IEEE* 58 (12): 1950–1951.

70 Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65 (6): 386.

71 Hey, T. (1999). Richard Feynman and Computation. *Contemporary Physics* 40 (4): 257–265.

72 Mead, C. and Ismail, M. (2012). *Analog VLSI Implementation of Neural Systems*, vol. 80. Springer Science & Business Media.

73 Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275 (5297): 213–215.

74 Markram, H., Gerstner, W., and Sjöström, P.J. (2011). A history of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience* 3 (4): 1–24.

75 Bi, G.Q. and Poo, M.M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* 18 (24): 10464–10472.

76 Morrison, A., Diesmann, M., and Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics* 98 (6): 459–478.

77 Raoux, S., Burr, G.W., Breitwisch, M.J. et al. (2008). Phase-change random access memory: a scalable technology. *IBM Journal of Research and Development* 52 (4.5): 465–479.

78 Burr, G.W., Brightsky, M.J., Sebastian, A. et al. (2016). Recent progress in phase-change memory technology. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 6 (2): 146–162.

79 Suri, M., Bichler, O., Querlioz, D. et al. (2012). Physical aspects of low power synapses based on phase change memory devices. *Journal of Applied Physics* 112 (5): 054904.

80 Eryilmaz, S.B., Kuzum, D., Jeyasingh, R.G. et al. (2013). Experimental demonstration of array-level learning with phase change synaptic devices. In: 2013 IEEE International Electron Devices Meeting (IEDM), pp. 621–624. IEEE.

81 Valov, I., Waser, R., Jameson, J.R., and Kozicki, M.N. (2011). Electrochemical metallization memories—fundamentals, applications, prospects. *Nanotechnology* 22 (25): 254003.

82 Wong, H.S.P., Lee, H.Y., Yu, S. et al. (2012). Metal–oxide RRAM. *Proceedings of the IEEE* 100 (6): 1951–1970.

83 Burr, G.W., Shelby, R.M., Sebastian, A. et al. (2017). Neuromorphic computing using non-volatile memory. *Advances in Physics: X* 2 (1): 89–124.

84 Eryilmaz, S.B., Kuzum, D., Yu, S., and Wong, H.S.P. (2015). Device and system level design considerations for analog-non-volatile-memory based neuromorphic architectures. In: 2015 IEEE International Electron Devices Meeting (IEDM), pp. 64–67. IEEE.

85 Suri, M., Bichler, O., Hubert, Q. et al. (2012). Interface engineering of pcm for improved synaptic performance in neuromorphic systems. In: 2012 4th IEEE International Memory Workshop (IMW), pp. 1–4. IEEE.

86 Suri, M., Bichler, O., Hubert, Q. et al. (2013). Addition of HfO_2 interface layer for improved synaptic performance of phase change memory (PCM) devices. *Solid-State Electronics* 79: 227–232.

87 Ambrogio, S., Ciocchini, N., Laudato, M. et al. (2016). Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Frontiers in Neuroscience* 10 (56): 1–12.

88 Jackson, B.L., Rajendran, B., Corrado, G.S. et al. (2013). Nanoscale electronic synapses using phase change devices. *ACM Journal on Emerging Technologies in Computing Systems (JETC)* 9 (2): 12.

89 Ohno, T., Hasegawa, T., Nayak, A. et al. (2011). Sensory and short-term memory formations observed in a Ag_2S gap-type atomic switch. *Applied Physics Letters* 99 (20): 203108.

90 Ohno, T., Hasegawa, T., Tsuruoka, T. et al. (2011). Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nature Materials* 10 (8): 591–595.

91 Jo, S.H., Chang, T., Ebong, I. et al. (2010). Nanoscale memristor device as synapse in neuromorphic systems. *Nano Letters* 10 (4): 1297–1301.

92 Suri, M., Bichler, O., Querlioz, D. et al. (2012). CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: auditory (cochlea) and visual (retina) cognitive processing applications. In: 2012 IEEE International Electron Devices Meeting (IEDM), pp. 10–13. IEEE.

93 Ziegler, M., Soni, R., Patelczyk, T. et al. (2012). An electronic version of Pavlov's dog. *Advanced Functional Materials* 22 (13): 2744–2749.

94 Sillin, H.O., Aguilera, R., Shieh, H.H. et al. (2013). A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology* 24 (38): 384004.

95 Xu, Z., Mohanty, A., Chen, P.Y. et al. (2014). Parallel programming of resistive cross-point array for synaptic plasticity. *Procedia Computer Science* 41: 126–133.

96 Orchard, G., Lagorce, X., Posch, C. et al. (2015). Real-time event-driven spiking neural network object recognition on the spinnaker platform. In: 2015 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2413–2416. IEEE.

97 Sheridan, P.M., Cai, F., Du, C. et al. (2017). Sparse coding with memristor networks. *Nature Nanotechnology* 12: 784–789.

98 Choi, H., Jung, H., Lee, J. et al. (2009). An electrically modifiable synapse array of resistive switching memory. *Nanotechnology* 20 (34): 345201.

99 Yu, S., Wu, Y., Jeyasingh, R. et al. (2011). An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation. *IEEE Transactions on Electron Devices* 58 (8): 2729–2737.

100 Wu, Y., Yu, S., Wong, H.S.P. et al. (2012). AlO_x-based resistive switching device with gradual resistance modulation for neuromorphic device application. In: 2012 4th IEEE International Memory Workshop (IMW), pp. 1–4. IEEE.

101 Yu, S., Gao, B., Fang, Z. et al. (2013). A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation. *Advanced Materials* 25 (12): 1774–1779.

102 Yu, S., Gao, B., Fang, Z. et al. (2012). A neuromorphic visual system using RRAM synaptic devices with Sub-pJ energy and tolerance to variability: Experimental characterization and large-scale modeling. In: 2012 IEEE International Electron Devices Meeting (IEDM), pp. 10–14. IEEE.

103 Piccolboni, G., Molas, G., Portal, J.M. et al. (2015). Investigation of the potentialities of Vertical Resistive RAM (VRRAM) for neuromorphic applications. In: 2015 IEEE International Electron Devices Meeting (IEDM), pp. 447–450. IEEE.

104 Merolla, P.A., Arthur, J.V., Alvarez-Icaza, R. et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345 (6197): 668–673.

105 Deutsch, D. (1985, July). Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 400 (1818): 97–117, The Royal Society.

106 Shor, P.W. (1999). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review* 41 (2): 303–332.

107 Grover, L.K. (1997). Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters* 79 (2): 325.

108 Veldhorst, M. et al. (2014). An addressable quantum dot qubit with fault-tolerant fidelity. *Nature Nanotechnology* 9: 981–985.

109 Itoh, K.M. and Watanabe, H. (2014). Isotope engineering of silicon and diamond for quantum computing and sensing applications. *MRS Communications* 4 (4): 143–157.

110 Maune, B.M., Borselli, M.G., Huang, B. et al. (2012). Coherent singlet-triplet oscillations in a silicon-based double quantum dot. *Nature* 481 (7381): 344–347.

111 DiVincenzo, D.P. (2000). *The Physical Implementation of Quantum Computation*. Wiley-VCH. *arXiv preprint quant-ph/0002077*.

112 Nielsen, M.A. and Chuang, I.L. (2000). *Quantum computation and Quantum Information*. Cambridge: Cambridge University Press.

113 Kok, P., Munro, W.J., Nemoto, K. et al. (2007). Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics* 79 (1): 135.

114 Brown, K.R., Wilson, A.C., Colombe, Y. et al. (2011). Single-qubit-gate error below 10− 4 in a trapped ion. *Physical Review A* 84 (3): 030303.

115 Waldherr, G., Wang, Y., Zaiser, S. et al. (2014). Quantum error correction in a solid-state hybrid spin register. *Nature* 506 (7487): 204–207.

116 Dolde, F., Bergholm, V., Wang, Y. et al. (2014). High-fidelity spin entanglement using optimal control. *Nature communications* 5: 3371.

117 Muhonen, J.T., Dehollain, J.P., Laucht, A. et al. (2014). Storing quantum information for 30 seconds in a nanoelectronic device. *Nature nanotechnology* 9 (12): 986–991.

118 Loss, D. and DiVincenzo, D.P. (1998). Quantum computation with quantum dots. *Physical Review A* 57 (1): 120.

119 Koppens, F.H.L., Buizert, C., Tielrooij, K.J. et al. (2006). Driven coherent oscillations of a single electron spin in a quantum dot. *Nature* 442 (7104): 766–771.

120 Petta, J.R., Johnson, A.C., Taylor, J.M. et al. (2005). Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* 309 (5744): 2180–2184.

121 Veldhorst, M., Yang, C.H., Hwang, J.C.C. et al. (2015). A two-qubit logic gate in silicon. *Nature* 526 (7573): 410–414.

122 Watson, T.F., Weber, B., Hsueh, Y.L. et al. (2017). Atomically engineered electron spin lifetimes of 30 s in silicon. *Science Advances* 3 (3): e1602811.

123 Devoret, M.H. and Schoelkopf, R.J. (2013). Superconducting circuits for quantum information: an outlook. *Science* 339 (6124): 1169–1174.

124 Vion, D., Aassime, A., Cottet, A. et al. (2002). Manipulating the quantum state of an electrical circuit. *Science* 296 (5569): 886–889.

125 Manucharyan, V.E., Koch, J., Glazman, L.I., and Devoret, M.H. (2009). Fluxonium: single cooper-pair circuit free of charge offsets. *Science* 326 (5949): 113–116.

126 Steffen, M., Kumar, S., DiVincenzo, D.P. et al. (2010). High-coherence hybrid superconducting qubit. *Physical Review Letters* 105 (10): 100502.

127 Vandersypen, L. and van Leeuwenhoek, A. (2017). 1.4 Quantum computing-the next challenge in circuit and system design. In: 2017 IEEE International Solid-State Circuits Conference (ISSCC), pp. 24–29. IEEE.