ABSTRACT
The world's first microprocessor, the 4004, was co-developed by Busicom, a Japanese manufacturer of calculators, and Intel, a U.S. manufacturer of semiconductors. The basic architecture of 4004 was developed in August 1969; a concrete plan for the 4004 system was finalized in December 1969; and the first microprocessor was successfully developed in March 1971. Microprocessors, which became the "technology to open up a new era," brought two outstanding impacts, "power of intelligence" and "power of computing". First, microprocessors opened up a new "era of programming" through replacing with software, the hardwired logic based on IC's of the former "era of logic". At the same time, microprocessors allowed young engineers access to "power of computing" for the creative development of personal computers and computer games, which in turn led to growth in the software industry, and paved the way to the development of high-performance microprocessors. In this paper, the birth, evolution and future of Microprocessor will be described. 4004 performance was only 0.06MIPS with 2,238 transistors and 750 KHz operating frequency. Microprocessors evolved from 4 bit to 64 bit microprocessors, introducing computer technologies such as pipeline, super-pipeline, super-scalar, VLIW, cache memory, and virtual memory system. Now, it is possible to integrate 16 sets of microprocessor with 64GB of memory on the board. In 20th century, microprocessors were used for increasing "power of intelligence". In 21st century, microprocessors will evolve into "tool to bring forth wisdom" for all mankind.

1. INTRODUCTION
The microprocessor, which evolved from the inventions of the transistor and the integrated circuit, is one of the icons of the present information age. The pervasiveness of the microprocessor in this age goes far beyond the wildest imagination at the time of the first microprocessor. From the fastest computers to the simplest toys, the microprocessor is found everywhere. The microprocessor today has over 1 billion transistors on some of the most powerful devices.

A microprocessor is a tiny, enormously powerful high speed electronic brain etched on a single silicon semiconductor chip which contains the basic logic, storage and arithmetic functions of a computer. It thinks for the computer and, like a traffic cop, coordinates its operations. It receives and decodes instructions from input devices like keyboards, disks then sends them over a bus system consisting of microscopic etched conductive "wiring" to be processed by its arithmetic calculator and logic unit. The results are temporarily stored in memory cells and released in a timed sequence through the bus system to output devices such as CRT Screens, networks, or printers. In some cases, the terms 'CPU' and 'microprocessor' are used interchangeably to denote the same device.

The different ways in which microprocessors are categorized are:

a) CISC (Complex Instruction Set Computers)
b) RISC (Reduced Instruction Set Computers)
a) VLIW(Very Long Instruction Word Computers)
b) Super scalar processors

2. BIRTH OF THE MICROPROCESSOR
In 1970, Intel introduced the first dynamic RAM, which increased IC memory by a factor of four. These early products identified Intel as an innovative young company. However, their next product, the microprocessor, was more than successful by setting in motion an engineering feat that dramatically altered the course of electronics.

The project that produced the microprocessor originated in 1969, when Nippon Calculating Machine Corporation approached Intel to design 12 custom chips for its new Busicom 141-PF printing calculator. Ted Hoff, the Intel engineer assigned to the project, believed the design was not cost effective. His solution was to simplify the design and produce a programmable processor capable of creating a set of complex special-purpose calculator chips.

Together with Federico Fagin, later the founder of Zilog, Hoff came up with a four-chip design; a ROM for custom application programs, a RAM for processing data, an I/O device, and an unnamed 4-bit central processing unit which would become known as a "microprocessor."

Busicom, initially skeptical, finally gave Intel the go ahead. Nine months later, Intel successfully completed the project. Remarkably, the four-bit 4004 microprocessor, composed of 2,300 transistors etched on a tiny chip, could execute 60,000 instructions per second, making it as powerful as the massive ENIAC, but despite the Busicom success, Intel itself was divided over the microprocessor's broader application. Intel negotiated a return of the Busicom design rights and committed to developing the microprocessor's potential.

The first fully functional 4004 parts were available in March 1971, with the first public announcement in November 1971;
Intel 4004 was the first commercially available single-chip microprocessor in history. It was a 4-bit CPU designed for usage in calculators, designed for "embedded applications". Clocked at 740 KHz, the 4004 executed up to 92,000 single word instructions per second, could access 4 KB of program memory and 640 bytes of RAM. The Intel 4004 was a part of MCS-4 chipset, which included the following chips:

i. 4001 - 256-bit mask ROM and 4-bit I/O device,
ii. 4002 - 320-bit RAM and 4-bit I/O device,
iii. 4003 - 10-bit shift register,
iv. 4008 and 4009 - standard memory and I/O interface set.

**Figure 1. Intel's first 4004 microprocessor**

As the first single-chip microprocessor, the Intel 4004 is very popular with CPU collectors and non-collectors. Earlier Intel C4004 CPUs in white ceramic package are sought-after by beginner and intermediate collectors, and are usually sold for hundreds of dollars. The only known second source manufacturer of 4004 microprocessors was National Semiconductor.

3. EVOLUTION

In the earliest stages, microprocessors mainly filled the needs of embedded applications. As we trace the history of microprocessor, we will explore its evolution and the driving forces behind this evolution. The evolution of microprocessors can be divided into 5 generations.

3.1 First Generation (1971-73)

The microprocessors that were introduced from 1971 to 1973 were referred to as the first generation systems. First-generation microprocessors processed their instructions serially i.e., they fetched the instruction, decoded it, and then executed it. When the instruction was completed, the microprocessor updated the instruction pointer and fetched the next instruction, performing this sequential drill for each instruction in turn.

The Intel 4004 is a complete 4-bit parallel central processing unit (CPU). The 4004 easily interfaces with keyboards, switches, displays, AD converters, printers, and other peripheral equipment. The CPU can directly address 4K 8-bit instruction words of program memory and 5120 bits of data storage RAM. 16 Index registers are provided for temporary data storage. Up to 16 4-bit input ports and 16 4-bit output ports may also be directly addressed. The 4004 is fabricated with P-channel silicon gate MOS technology. Below is the block diagram of the 4004.

**Figure 2. Block diagram of 4004**

The Intel 4004 was followed in 1972 by the Intel 8008, the world's first 8-bit microprocessor. The 8008 is not an extension of the 4004 design, but the culmination of a separate design project at Intel, arising from a contract with Computer Terminals Corporation of San Antonio TX, for a chip they were designing. In 1969, CTC contracted two companies, Intel and Texas Instruments, to make a single-chip implementation, known as the CTC 1201. In late 1970, Texas Instruments dropped out being unable to make reliable parts. In 1970, with Intel yet to deliver the part, CTC opted to use their own implementation in the Data point 2200, using traditional TTL logic instead. Intel's version of the 1201 microprocessor arrived in late 1971, but was too late, slow, and required a number of additional support chips. CTC had no interest in using it. CTC had originally contracted Intel for the chip, and would have owed them $50,000 for their design work. To avoid paying for a chip they did not want, CTC released Intel from their contract and allowed them free use of the design. Intel marketed it as the 8008 in April, 1972, as the world's first 8-bit microprocessor. It was the basis for the famous "Mark-8" computer kit advertised in the magazine Radio-Electronics in 1974.

There were other microprocessors in the market during the same period:

i. Rockwell International's PPS-4 (4 bits)
ii. National Semiconductor's IMP-16 (16 bits)

They were fabricated using PMOS technology which provided low cost, slow speed and low output currents. They were not compatible with TTL.

3.2 Second Generation (1974-78)

By the late 1970s, enough transistors were available on the IC to usher in the second generation of microprocessor sophistication. The second generation marked the beginning of very efficient 8 bit microprocessors. Some of the popular processors were:

i. Intel's 8080
ii. Motorola's 6800 and 6809
iii. Zilog's Z80
This generation is defined by overlapped fetch, decode, and execute steps (Computer 1996). As the first instruction is processed in the execution unit, the second instruction is decoded and the third instruction is fetched. The distinction between the first and second generation devices was primarily the use of newer semiconductor technology to fabricate the chips. This new technology resulted in a five-fold increase in instruction, execution, speed, and higher chip densities. They were manufactured using NMOS technology. This technology offered faster speed and higher density than PMOS. It is TTL compatible.

3.2.1 Intel 8080
Intel’s experience with the 8008 provided a tremendous source of ideas on how to improve on the microprocessor. Starting in the middle of 1972, these ideas were used to define the Intel 8080 microprocessor. The improvements in the 8080 included more instructions, a 64-KB address space, 256 I/O ports, 16-bit arithmetic instructions, and vectored interrupts. The designers of the 8080 included some of the key individuals responsible for the 4004 and 8008, Federico Faggin and Masatoshi Shima. The 8080 was introduced in early 1974 with a price tag of $360. The 8080 was designed in 6-micron MOS with n-type transistor (NMOS) technology and required 6,000 transistors. The 40-pin package allowed for separate address and data buses. The first 8080 ran at 2 MHz and was rated at 0.64 millions of instructions per second (MIPS). Unlike the 4004 and 8008, the 8080 was quickly adopted by designers. It was incorporated into numerous products, the most significant being the Altair 8800 microcomputer kit from a company called MITS. The Altair showed there was a market for microprocessors beyond traditional embedded applications.

3.2.2 Motorola’s 6800
Motorola entered the microprocessor market in 1974 with the 8-bit 6800. The 6800 required 4,000 transistors and was fabricated in NMOS technology. The 6800 offered some significant benefits over the 8080, including improved performance and the need for only a single 5-volt supply. The 6800 contained two 8-bit general-purpose registers and a single index register, which meant that it operated on data primarily in memory. Because the memory technology at the time was faster than the microprocessor, accessing memory did not impose a performance penalty. Motorola first produced a custom version of the 6800 for General Motors and later for Ford. This was the beginning of a huge market for embedded processors in cars, which Motorola has since dominated. Variants of the 6800 have been introduced over the years, including the 6809 in 1977, the 6801, the 68HC11, and the 68HC16.

3.2.3 Zilog’s Z80
Faggin and Shima left Intel in 1975 to form Zilog, which would produce the Z80. The 2.5-MHz Zilog Z80 was released in 1976 and offered compatibility with the 8080, along with many significant enhancements. The instruction set was expanded and included block move and block I/O instructions. A second register set was added to better support interrupts and operating systems (OSs). The Z80 interface simplified the system design by providing dynamic random access memory (DRAM) refresh signals and an on-chip clock circuit, which could be connected directly to an external crystal. The Z80 would outsell the 8080 as it became the microprocessor of choice in many applications. The most significant microcomputer application, the Tandy TRS-80, was introduced in 1977. The TRS-80 contained a Z80, 4-KB RAM, 4-KB ROM, a keyboard, a black and white video display, and a tape cassette, all for $600. Thousands were sold in the first few months, exceeding all projections. To this day, the Z80 continues to be a popular microprocessor in embedded applications.

Figure 3. Block diagram of Z80

3.2.4 Other Noteworthy Microprocessors
The 8-bit RCA 1802, introduced in 1974, was one of the first microprocessors designed using complementary MOS (CMOS) technology. The 1802 ran at 6.4 MHz with a 10-volt supply, making it one of the fastest microprocessors of its time. Its simple design included sixteen 16-bit registers, which were also usable as thirty-two 8-bit registers. It used an 8-bit opcode to implement the limited instruction set. The most significant applications of the 1802 were in several NASA space probes. It was used in those cases because a version that used the radiation-resistant silicon-on-sapphire technology was available.

The 8-bit National Semiconductor single-chip microprocessor (SC/MP), introduced in 1976, was the first microprocessor to support multiple bus masters on its system bus. This feature supported multiple SC/MPs and other bus masters, such as a direct memory access (DMA) controller. Arbitration was controlled by a “daisy chain” connecting the bus masters in priority order. The EOUT (enable out) and ENIN (enable in) signals of the SC/MP were used to chain the processors together. Another unique feature of the SC/MP was its bit serial arithmetic logic unit (ALU).

The 16-bit TI TMS9900, introduced in 1976, was the first single-chip 16-bit microprocessor. Its architecture was based on the TI 990 minicomputer. The TMS9900 had only two 16-bit internal registers, with one of them pointing to the memory-resident register set. The speed of memory at the time made it feasible to use external memory for the register set. A simple adjustment of the internal register could be used to save the registers for a procedure call or interrupt. A version of the TMS9900, the TMS9940, was used in the TI 99/4 PC, introduced in 1979.
3.3 Third Generation (1979-80)

The third generation, introduced in 1979, was represented by Intel’s 8086 and the Zilog Z8000, which were 16 bit processors with minicomputer-like performance. The third generation came about as IC transistor counts approached 250,000. Motorola’s MC68020, for example, incorporated an on-chip cache for the first time and the depth of the pipeline increased to five or more stages. This generation of microprocessors was different from the previous ones in that all major workstation manufacturers began developing their own RISC-based microprocessor architectures (Computer, 1996). This age is dominated by 16 - bit microprocessors

3.3.1 Intel’s 8086

The 8086 microprocessor was structured as a bus interface unit (BIU) and an execution unit (EU). The BIU handled instruction and operand fetches from memory. The BIU fed opcodes to and requested operands from the EU, which performed the instructions. The BIU and EU constituted a simple pipeline, with the BIU fetching instructions concurrently with processing in the EU. The 8086 was source-code-compatible with the 8080/8085. It used variable-length instructions of one or more bytes fetched into the pre fetch queue. The four 16-bit registers could be used as either 16-bit or 8-bit registers. The 8086 instituted an unusual form of segmented addressing. Within a segment, addressing was limited to 64 KB. Addressing was expanded to 1 MB by the addition of the segment register shifted by four to a 16-bit address. The 80286 extended addressing to 16 MB, but still through segments of no more than 64 KB and only in “protected” mode as opposed to the 8086’s “real” mode. The 8086 had a companion floating-point chip, the 8087. The 8087 introduced Intel’s 80-bit floating-point format, greatly influencing the IEEE floating-point standard 754, issued in 1985.

![Block diagram of 8086](image)

3.3.2 Motorola’s 68000

The 68000 had a more orthogonal architecture than the 8086. The 68000 fetched instructions of one or more 16-bit words. It featured 32-bit address and data registers, providing a linear address space and a path to future full 32-bit implementations. The 68000 had a simple pipeline, overlapping instruction fetch and execution. The 68010 added virtual memory support through the ability to restart instructions on a page fault. The 68020 was one of the first true 32-bit processors with a true pipeline, overlapping operand access with internal execution. It also was one of the first microprocessors with an on-chip instruction cache of 256 bytes.

3.3.3 Zilog’s Z8000

The Z8000 was Zilog’s follow-on to the successful Z80. However, the Z8000 sacrificed the compatibility of the Z80 to make better use of a 16-bit external bus to memory and to make the instruction set orthogonal with respect to its 16 general-purpose registers. The Z80’s 8-bit opcodes could not encode more than one of the 16 registers as an operand. The Z8000’s 16-bit registers could also be used as thirty-two 8-bit registers, eight 32-bit registers, and even as four 64-bit registers. The Z8000 was not pipelined because it was felt that the fixed 16-bit instruction format and simple address calculation eliminated the need for prefetching. The Z8000 was also singular in using hardwired logic instead of microcode ROM, in spite of increasing the instruction set from 128 instructions in the Z80 to 414. This may have contributed to its lack of success, since it suffered from initial bugs.

3.4 Fourth Generation (1981-1995)

As the workstation companies converted from commercial microprocessors to in-house designs, microprocessors entered their fourth generation with designs surpassing a million transistors. This era marked the beginning of 32 bits microprocessors. Intel introduced 432, which was bit problematic and then Intel 80386 was launched. Motorola introduced 68020/68030.

They were fabricated using low-power version of the HMOS technology called HCMOS. Motorola introduced 32-bit RISC processors called MC68100.

Intel's first 32-bit microprocessor was the iAPX 432, which was introduced in 1981 but was not a commercial success. It had an advanced capability-based object-oriented architecture, but poor performance compared to contemporary architectures such as Intel's own 80286 (introduced 1982), which was almost four times as fast on typical benchmark tests.

Motorola’s success with the 68000 led to the MC68010, which added virtual memory support. The MC68020, introduced in 1984 added full 32-bit data and address buses. The 68020 became hugely popular in the UNIX super microcomputer market, and many small companies produced desktop-size systems. The MC68030 was introduced next, improving upon the previous design by integrating the MMU into the chip. The continued success led to the MC68040, which included an FPU for better math performance. A 68050 failed to achieve its performance goals and was not released, and the follow-up MC68060 was released into a market saturated by much faster RISC designs. The 68k family faded from the desktop in the early 1990s.

Other large companies designed the 68020 and follow-ons into embedded equipment. At one point, there were more 68020s in embedded equipment than there were Intel Pentiums in PCs

3.4.1 Alpha and PowerPC

The Alpha21064 and PowerPC601 best illustrate the contrasting designs of the various RISC CPUs and are considered below in
some detail. Both processors were load-store architectures, with 32-bit instructions and two 32/64-bit register files for floating point and integer. The Alpha designers focused on very fast clocks, a simple instruction set that would enable fast clocking, and deep pipelines. The PowerPC instruction set had powerful instructions that did more in each clock. Of the three factors that affect performance, Alpha chose to reduce the clock period and CPI at the expense of the number of instructions. The PowerPC601 took a more balanced approach. The clock rate of a CPU depends on the amount of logic in each pipeline stage. Thus, longer pipelines reduce the amount of logic in each stage and allow faster clocks. Unfortunately, branches in program execution cause greater penalties in deep pipelines. Therefore, prediction of branches has become critical to high performance.

During this time, National Semiconductor introduced a very similar 16-bit pin out, 32-bit internal microprocessor called the NS 16032, the full 32-bit version named the NS 32032. Later, National Semiconductor produced the NS 32132, which allowed two CPUs to reside on the same memory bus with built in arbitration. The appearance of RISC processors like the AM29000 and MC88000 which no more exist now influenced the architecture of the final core, the NS32764. From 1993 to 2003, the 32-bit x86 architectures became increasingly dominant in desktop, laptop, and server markets and these microprocessors became faster and more capable.

3.5 Fifth Generation (1995- Till date)

Microprocessors in their fifth generation, employed decoupled super scalar processing, and their design soon surpassed 10 million transistors. In this generation, PCs are a low-margin, high-volume-business dominated by a single microprocessor. This age the emphasis is on introducing chips that carry on-chip functionalities and improvements in the speed of memory and I/O devices along with introduction of 64-bit microprocessors.

Intel leads the show here with Pentium, Celeron and very recently dual and quad core processors working with up to 3.5GHz speed.

While 64-bit microprocessor designs have been in use in several markets since the early 1990s the early 2000s saw the introduction of 64-bit microprocessors targeted at the PC market. With AMD’s introduction of a 64-bit architecture backwards-compatible with x86, x86-64 followed by Intel’s near fully compatible 64-bit extensions, the desktop era began. Both versions can run 32-bit legacy applications without any performance penalty as well as new 64-bit software. With operating systems Windows XP x64, Windows Vista x64, Windows 7 x64, Linux, BSD, and Mac OS X that run 64-bit native, the software is also geared to fully utilize the capabilities of such processors. The move to 64 bits is more than just an increase in register size from the IA-32 as it also doubles the number of general-purpose registers.

The move to 64 bits by PowerPC processors had been intended since the processors’ design in the early 90s and was not a major cause of incompatibility. Existing integer registers are extended as are all related data pathways, but, as was the case with IA-32, both floating point and vector units had been operating at or above 64 bits for several years. In 2011, ARM introduced a new 64-bit ARM architecture. The latest development is the introduction of dual-core processors from both Intel and AMD. Dual-core processors have two full CPU cores operating off of one CPU package in essence enabling a single processor to perform the work of two processors.

3.5.1 MIPS and HP

At its introduction in 1992, the MIPS R4000 was one of the fastest single-chip processors, with a super pipelined 64-bit architecture. This architecture was engendered by the high-end graphics market that Silicon Graphics dominated. The external clock (50 MHz) was doubled in the CPU to clock the deep pipelines at 100 MHz. Address and data buses were 64-bit and multiplexed. The R4000 had separate direct-mapped instruction and data caches of 8 KB and a second-level cache controller on chip. Several variations of the R4000 were made in the following years and the MIPS architecture became popular in the embedded marketplace.

HP manufactured its own processors for its workstations. Therefore, PA-RISC was more proprietary than MIPS or Sparc. The PA-RISC 7100 and 7200 were 32-bit processors with external cache that required systems to use very-high speed static RAMs (SRAMs). The 7200, produced in 1994, had two integer units and one FPU. It dispatched two instructions to any of the three units. The first 64-bit architecture from HP was the 180-MHz PA-RISC 8000, produced in 1996. For a short time it surpassed Digital’s 333-MHz Alpha 21164 in integer performance. The pattern for most of the 1990s has been that every new processor introduced tends to surpass its older rivals. The only exception has been Alpha, which has held the top spot for most of the decade. The threat to these traditional RISC vendors is the proliferation of x86 and the encroachment of Windows NT into the UNIX market.

4. FUTURE OF MICROPROCESSOR

Considering the dramatic progress made in microprocessor design and architecture since 1965, it’s risky to project what new technologies could become available in ten to fifteen years. And yet, a group of future-minded researchers are expressing optimism about the potential of tiny Nano electronic components, organic molecules, carbon nanotubes and individual electrons that could serve as the underlying technology for new generation of microprocessors emerging around 2015. There are limits to what can be accomplished with silicon, says Philip J. Kuekes, a physics researcher at Hewlett-Packard Laboratories quoted in The New York Times. Kuekes says H-P is currently working on molecular-scale nanotechnology switches that, it is hoped, will be able to overcome some of the present-day technological challenges.

Kuekes, along with electrical engineers, chemists and physicists from around the world, are collaborating with various semiconductor manufacturers and suppliers, government organizations, consortia and universities to promote advancements in the performance of microprocessors and solve some of the challenges that cast doubt on the continuation of the advancements described by Moore’s Law.

The mid-term future of microprocessor advancements could very well be based on nanotechnology designs that overcome the physical and quantum problems associated with conventional silicon transistors and processor cores.

Microprocessor technology has delivered three-orders-of-magnitude performance improvement over the past two decades, so continuing this trajectory would require at least 30x performance increase by 2020. Microprocessor-performance scaling faces new challenges precluding use of energy-inefficient
microarchitecture innovations developed over the past two decades. Further, chip architects must face these challenges with an ongoing industry expectation of a 30x performance increase in the next decade and 1,000x increases by 2030.

As the transistor scales, supply voltage scales down, and the threshold voltage of the transistor (when the transistor starts conducting) also scales down. But the transistor is not a perfect switch, leaking some small amount of current when turned off, increasing exponentially with reduction in the threshold voltage. In addition, the exponentially increasing transistor-integration capacity exacerbates the effect; as a result, a substantial portion of power consumption is due to leakage. To keep leakage under control, the threshold voltage cannot be lowered further and, indeed, must increase, reducing transistor performance.

As transistors have reached atomic dimensions, lithography and variability pose further scaling challenges, affecting supply-voltage scaling. With limited supply-voltage scaling, energy and power reduction is limited, adversely affecting further integration of transistors. Therefore, transistor-integration capacity will continue with scaling, though with limited performance and power benefit. The challenge for chip architects is to use this integration capacity to continue to improve performance.

5. CONCLUSION

The future looks bright for Microprocessor with the advancement in technology and it is constantly improving every year. We will soon be able to see more miniature devices that are powerful, energy efficient and surely blow your mind. On the other side, microprocessor hardware improvements are becoming more and more difficult to accomplish as, even Gordon Moore believes, the exponential upward curve in microprocessor hardware advancements “can’t continue forever.” Because the future winners are far from clear today, it is way too early to predict whether some form of scaling (perhaps energy) will continue or there will be no scaling at all. Moreover, the challenges processor design will face in the next decade will be dwarfed by the challenges posed by these alternative technologies, rendering today's challenges a warm-up exercise for what lies ahead.

6. REFERENCES


