

纳米级工艺对微处理器设计的挑战

胡伟武 李国杰

摘要：随着集成电路制造技术进入纳米级，国际上高性能通用 CPU 的发展正面临技术转型期。性能功耗比继性能价格比之后正在成为计算机的主要设计指标，主频至上的计算机处理器设计技术正在终结，互连网的普及正在改变计算机的应用模式，EDA 工具不断完善和成熟以及集成电路代工厂正在蓬勃兴起。上述技术转型为我国在未来几年发挥后发优势，另辟蹊径，通过跨越创新实现突破提供了机遇。我们应该紧紧抓住上述机遇，加大研发和产业化力度，实现跨越发展。

关键词：微处理器；摩尔定律；系统结构；技术转型；跨越创新

1 引言

在计算机发明以来的几十年中，半导体工艺技术和计算机系统设计技术互为动力、互相促进推动着计算机工业乃至信息产业的蓬勃发展。一方面，半导体工艺水平的提高为计算机系统的设计者提供了更多更快的晶体管来实现具有更多功能、更高性能的系统；另一方面，计算机系统技术的发展为工艺进步提供了强大的动力，世界上最先进的半导体工艺都用于生产计算机用的处理器芯片，这些工艺为处理器生产厂家所拥有。

应用需求也是计算机系统发展的动力源泉，但在工艺、系统和应用三者的关系中，应用需求对工艺和系统的拉动作用远小于工艺和系统的发展对应用的推动作用。在主流的 PC 市场，总是英特尔和微软从获取利润的角度引导计算机应用的发展；互联网也是先由科学家设想和开发，后来才逐步向大众普及；计算机系统设计人员经常问的问题是“如何充分利用工艺发展带来的数以亿计的晶体管”，而不是“满足特定应用的需求需要多少晶体管”。正是由于工艺技术的快速发展导致了计算机工业乃至整个信息产业在工艺、系统和应用之间出现上述“本末倒置”的现象。

20 世纪 60 年代，英特尔公司的创始人摩尔（Gordon Moore）预测集成电路中单位面积集成的晶体管数目大约每 18 个月翻一番。此后的三十多年，半导体工艺技术基本上按照摩尔定律的预测发展。目前的微处理器已经在片内集成十几亿个晶体管。根据世界半导体行业共同制订的 2005 年国际半导体技术发展路线图，未来 15 年集成电路仍将按摩尔定律持续高速发展。半导体技术现在已经允许将许多不同类型的晶体管（如 P 型和 N 型沟道 MOS 晶体管，PNP 和 NPN 双极性晶体管，浮动栅器件，熔丝和反熔丝）集成到同一个衬底上，并允许将处理器和动态存储器集成到一块芯片上。半导体技术的这些进步，为处理器的设计者提供了更多的资源来实现更高性能的芯片，从而有可能在单个芯片上创造更复杂和更灵活的系统。

然而，随着工艺特征尺寸（feature size）缩小到纳米级，工艺技术对结构的影响通过几十年的积累产生了质的变化，主要体现在以下几个方面：

- (1) 片内晶体管数目的增加大大增加了芯片复杂度(目前复杂的微处理器内部已经集成了几十亿个晶体管)，晶体管特征尺寸的缩小则增加了物理设计的难度（纳米级的物理设计需要考虑串扰、片内参数漂移、可生产性、电源完整性等一系列问题），这些都大幅度增加了设计成本及设计周期；

- (2) 在 0.13 微米之后晶体管工作电压难以随着工艺的进步而降低，虽然每个晶体管的功耗随着特征尺寸的缩小有所减少，但晶体管数目的增加以及主频的提高使得整个芯片的功耗大幅度增加。此外纳米级工艺中晶体管的漏电量大幅度增加更对功耗增加起着推波助澜的作用。功耗问题已经成为芯片系统设计的主要瓶颈。如果沿用目前的电路和结构，到 2018 年左右，微处理器芯片的功耗将超过封装功耗极限（ $200\text{W}/\text{mm}^2$ ）的 4 倍（即达到 $1\text{KW}/\text{mm}^2$ ）；
- (3) 集成度的提高意味着线宽变窄，信号在片内传输单位距离所需的延迟也相应增大，在现代的高性能微处理器中，信号在一个时钟周期内传输的距离只相当芯片尺寸的十分之一左右。导致连线延迟而不是晶体管翻转速度将越来越成为影响处理器主频的主要因素；
- (4) 随着片内处理能力的进一步加强，封装能力成为制约系统性能进一步提高的瓶颈，管脚数目限制了芯片对外通信能力，导致“茶壶里倒饺子”的不平衡设计。

为了解决这些问题，芯片设计越来越强调结构的层次化、功能部件的模块化和分布化，即每个功能部件都相对地简单，部件内部尽可能保持通信的局部性。目前，片内多核处理器及多线程技术较好地解决了上述问题，正在成为处理器体系结构设计的热点技术。

目前主流处理器中使用的多核结构同样也是由工艺技术驱动，而不是由应用需求拉动的。工艺技术的发展要求处理器采用分布的、并行的结构，而作为计算机理论基础的图灵机理论和作为计算机结构基础的冯·诺依曼结构本质是顺序的、集中的。目前还没有很好的方法解决并行程序编程困难、并行效率不高以及串行程序难以自动并行化等问题，多核结构只能用于粗粒度的任务级并行以及手工并行编程的场合。多核结构是不得已而为之，应用的首选仍是高主频低功耗设计。

由于上述工艺技术进步的影响，以及互联网的普及正在改变计算机的应用模式，因此目前计算机系统设计正面临着技术转型期。这为中国在未来几年发挥后发优势，另辟蹊径，通过跨越创新实现突破提供了机遇。

2 工艺与材料技术的发展

摩尔在 1965 年提出芯片的晶体管数量大约每年将翻一番，于 1975 年修正为每两年翻一番的预测。近四十年来，半导体工业的发展进程大致符合这项定律，工业界不断推出性能更高、功能更强大、单位成本更低的芯片，芯片的晶体管数量从数千个晶体管增加到数十亿个晶体管，工艺特征尺寸从几微米缩小到了 45 纳米。

根据世界半导体行业共同制订的 2005 年国际半导体技术发展路线图（ITRS）及其 2006 年更新，未来 15 年集成电路仍将按摩尔定律持续高速发展。ITRS 预测 2009 年将采用 45 纳米的工艺，2011 年将采用 32 纳米的工艺，2014 年将采用 22 纳米的工艺，2017 年将采用 16 纳米的工艺，2020 年将采用 11 纳米的工艺。ITRS 预测到 2013 年，高性能微处理器芯片上可集成的晶体管数将超过 88 亿个（到 2020 年超过 353 亿个），片上局部时钟频率将达到 22 GHz（到 2020 年达到 73GHz）。

在摩尔定律带来晶体管数量增加和性能提高的同时人们也发现，随着工艺发展到纳米级，晶体管的工作电压难以再线性降低，连线延迟而不是晶体管的开关延迟成为主要的电路延迟，漏电功耗随着晶体管阈值电压的降低而指数增长。处理器的时钟频率越来越难以提高，功耗问题越来越严重，系统散热也越来越难以解决，导致依传统途径扩展的空间十分有限。

在半导体工艺发展的头 35 年，场效应晶体管发展的努力集中在提高性能（增加器件的速度）以及集成更多的器件和功能到片上，但是最近的五年，器件特性日益明显的变化和芯片功耗密度持续的攀升成为主要的挑战。例如随着线宽尺寸的不断缩小，CMOS 器件机理适用范围面临着量子机制的边界，在 65 纳米工艺的时候，二氧化硅绝缘层的厚度已经降低至 1.2 纳米，约为 5 个硅原子层的厚度，隧道穿越引起的漏电电流急剧增加。

通过采用新技术和新工艺来克服这些困难可以继续延续摩尔定律。在 90/65 纳米制造工艺中，采用了多项新技术和新工艺，包括应变硅^[1]（Strained Silicon）、绝缘硅（或称绝缘层上覆硅，SOI- Silicon on Insulator）、铜互连技术、低 k 介电材料等。最近 45 纳米工艺所采用的高 k 介质和金属栅材料的技术是晶体管工艺技术的又一个重要突破。采用高 k 介质（SiO₂ 的 k 为 3.9，高 k 材料为 20 以上）如氧氮化铪硅（HfSiON），理论上相当于提升栅极的有效厚度，使漏电电流下降到 10% 以下，另外高 k 介质材料和现有的硅栅电极并不相容，必须采用新的金属栅电极材料来增加驱动电流。这种技术打开了通往特征尺寸为 32 纳米及 22 纳米的工艺通路，扫清了工艺技术中的一大障碍。摩尔称此举是 CMOS 工艺技术中的又一里程碑，将摩尔定律又延长了另一个 10 年至 15 年。

2000 年 12 月，英特尔（Intel）公司率先开发出栅极长度为 30 纳米的单晶体管；2001 年 11 月，英特尔宣布已开发出栅长仅为 15 纳米的新型晶体管，同时单个晶体管的实际工作频率已经达到了 2.63THz。英特尔发布的 15 纳米晶体管采用“耗尽型衬底晶体管（depleted substrate transistor）”的新型结构和绝缘硅技术及“高 k 栅电介质”材料，从而使制造出的芯片的晶体管数量可以达到现有微处理器的 25 倍，运行速度提高 10 倍。2002 年 12 月，IBM 宣布了当前世界上最细小的晶体管加工技术。利用该技术生产出的晶体管栅长仅为 6 纳米。能够以如此之小的尺寸制造出可实际动作的晶体管，意味着芯片的晶体管数量可以达到现有微处理器的 100 倍以上。

多栅晶体管技术是一种新型电路结构技术。传统晶体管是每个晶体管只有一个栅用来控制电流在两个结构单元之间通过或中断，进而形成计算中所需的“0”与“1”。而多栅晶体管技术是每个晶体管有两个或三个栅，从而提高了晶体管控制电流的能力（即计算能力），并降低了功耗，减少了电流间的相互干扰。目前，英特尔、AMD（超微半导体）和 IBM 公司已分别在实验室成功开发出多栅晶体管。2003 年 9 月，AMD 公布了采用全耗尽型绝缘硅（Fully-depleted SOI, FDSOI）、硅锗、三栅（Tri-Gate）和镍硅金属栅（NiSi）的栅长为 20 纳米的硅晶体管。IBM 则已开始致力于将双栅晶体管技术应用于芯片的生产，其硅锗生产工艺等方面的进展会加快双栅晶体管技术的产品化。英特尔于 2003 年 6 月在实验室实现了栅长为 30 纳米的三栅晶体管，预计在 2010 年前后实现三栅晶体管技术的产品化，并逐渐使三栅晶体管成为未来生产出尺寸更小、处理性能更强的芯片的关键技术。

3D 芯片技术是 IBM 公司、Matrix 半导体公司等研发的未来芯片技术。在一块芯片的设计中，这种技术将晶体管封装成两层或三层以上。它通过充分利用立体空间，将数量成倍增加的晶体管封装进尺寸并未发生多大变化的一块芯片中，缩短了晶体管之间连接的金属导线的长度，有助于增强芯片的性能。

微处理器尺寸的缩小和晶体管数量的增加也依赖于刻蚀技术的进步。随着半导体技术的发展，目前使用的可见光光刻技术将很快达到物理极限。极远紫外线光刻（extreme ultraviolet lithography, EUVL）是其替代技术。由于超紫外线的波长短于可见光，因此，刻写的“线条”更细。目前，由英特尔、摩托罗拉、AMD、Micron、Infineon 等多家公司和实验室成立的联合机构已开发出用于极远紫外光刻的光掩模制造技术；另外浸入式技术与现有的光刻技术兼容性好，工艺和装备变动不大，也获得了人们的极大赞赏，ITRS 预测浸入式光刻技术

会继续着其既有的发展态势，是 2007 年达到 45 纳米、2011 年达到 32 纳米节点的关键技术，而实现 45 纳米 COMS 工艺主要依靠 193 纳米浸入式光刻技术。

在不断改进和发展晶体管技术的同时，业界也开始考虑后 CMOS 时代的技术措施并积极寻找新的替代材料和器件，以便在特征尺寸更小的技术节点中超越硅 CMOS 技术。伴随着新的材料和器件结构的融入，半导体制造已经转向“材料时代”。ITRS 中提出的非传统 CMOS 器件包括超薄体绝缘硅、能带工程晶体管、垂直晶体管、双栅晶体管、FinFET^[2]等。未来有望被广泛应用的新兴存储器件主要有磁性存储器（MRAM）、纳米存储器（NRAM）、分子存储器（molecular memory）等。新兴的逻辑器件主要包括谐振隧道二极管、单电子晶体管器件、快速单通量子逻辑器件、量子单元自动控制器件、自旋电子器件（spintronic storage）、碳纳米管（carbon Nanotubes），硅纳米线（Silicon Nanowires）、分子电子器件（Molecular Electronics）等。

目前，AMD 提出用镍来制作栅极，因为镍栅能够增加通过的电流，减少电流泄露，而且可以做得比较厚，便于生产制造。而英特尔和 IBM 则致力于硅锗技术（SiGe，即应力硅技术）。硅锗技术通过在芯片的硅原子晶格之间填入锗，将晶体管中硅原子之间的距离进行了拉伸，从而使原子之间的相互作用力更小，这会极大地降低原子作用力对通过晶体管的电流的干扰，最终结果便是芯片处理性能的提高和功耗的降低。IBM 利用硅锗技术已加工出工作频率高达 350GHz 的晶体管，并用这种晶体管生产出频率约达 150GHz 的网络通信芯片。

在未来各种集成电路新兴器件中，大量运用了纳米技术。除了在存储器和逻辑器件中作为晶体管的主要材料，某些形态的碳纳米管可在晶体管中取代硅来控制电子流，还可取代铜作为互连材料。碳纳米管直径只有 1 纳米至 2 纳米，只是硅晶体管尺寸的 1/500。碳纳米管因其超常的能量及半导体性能而被认为是最有可能在未来取代硅，成为生产晶体管及微处理器的主要材料。此外，碳纳米管投入运行时产生的热量和功耗都比晶体管要小得多。IBM 科学家已经研制出世界上最小的计算机逻辑电路——一个由单分子碳组成的双晶体管元件。英特尔公司的技术人员曾预测将来芯片的晶体管结构将由碳纳米管或硅纳米导线构成。不过，碳纳米管还处在研究试验阶段，在产品中使用碳纳米技术的时间可能需要十年或更长。

在过去的 40 多年，半导体工业的发展突破了一个又一个看似不可能跨越的瓶颈，神奇地遵循着摩尔定律，这一切都得益于生产技术的不断进步。可以预见随着各种新工艺和新材料的采用，半导体工业的这种发展趋势仍将继续下去。

3 系统结构设计的发展趋势

3.1 技术发展趋势

未来体系结构的发展面临着以下因素带来的挑战。

1. 复杂度障碍

工艺技术的进步为系统设计者提供了更多的资源来实现更高性能的芯片，也导致了芯片设计复杂度的大幅度增加。一支现代处理器设计队伍动辄几百到几千人，但设计能力的增长还是远远赶不上复杂度提高的步伐，验证能力更是成为芯片设计的瓶颈。

另外，晶体管特征尺寸缩小到纳米级（现在 45nm 的芯片已经进入量产）对芯片的物理设计带来了巨大的挑战。纳米级芯片中连线尺寸缩小，相互间耦合电容所占比重加大，连线

间的信号串扰（crosstalk）日趋严重；硅片上的性能参数（如介电常数、掺杂浓度等）的漂移变化产生芯片内时钟树的偏差；由于晶体管尺寸的缩小使得刻蚀等过程难以处理，在芯片设计时就要充分考虑可制造性设计。

总之，随着工艺技术的发展，工艺所提供的晶体管更多了，也更“难用”了，导致设计周期和设计成本大幅度增加，为解决这些问题，对设计提出了结构层次化、功能模块化、布局分布化的要求。

2. 主频障碍。

摩尔定律本质上是预测晶体管的尺寸及翻转速度变化趋势的定律，但由于商业的原因，摩尔定律同时被赋予每 1.5 年处理器主频提高一倍的含义。事实上过去每代微处理器主频是上代产品的两倍，其中只有 1.4 倍来源于器件的按比例缩小，另外 1.4 倍来源于结构的优化，包括流水级中逻辑门数目的减少等技术。

但目前的高主频处理器中，指令流水线的划分已经很细了，一级流水级只有 10~15 级 FO4^[3]的延迟，考虑到控制流水线锁存器本身的延迟，实际留给有效处理工作的逻辑只有 6~9 级 FO4，已经难以再降低，而且不可能使用少于 6~8 个等效 4 扇出反相器产生出波形好的时钟脉冲。所以结构优化产生的主频提高潜力已经“消费”得差不多了。

另外芯片集成度的提高意味着线宽变窄，信号在片内传输单位距离所需的延迟也相应增大，制约处理器主频以及性能的提高的因素已逐渐从晶体管翻转速度转移到连线延迟。

3. 功耗障碍。

随着晶体管数目的增加以及主频的提高，功耗问题越来越突出。现代的通用处理器功耗峰值已经高达上百瓦。例如，Alpha 21364 为 100 瓦，AMD Opteron 是 95 瓦，英特尔的安腾 II 已超过 100 瓦。如果功耗超过 150 瓦，无论是芯片的封装还是主板的供电能力，都已经难以为继了。在移动计算领域，功耗更是压倒一切的指标。因此如何降低功耗的问题已经十分迫切。

CMOS 电路的功耗与主频和规模都成正比，与电压的平方成正比，而主频在一定程度上又跟电压成正比。降低功耗需要包括工艺技术、物理设计、体系结构设计、系统软件以及应用软件的共同努力。如果说传统的微处理器设计追求的是每秒运行的次数（performance per second）以及每一块钱所能买到的性能（performance per dollar），那么在今天，每瓦特功耗所得到的性能（performance per watt）已经成为越来越重要的指标。因此，必须探索新的结构，通过包括工艺技术、物理设计、体系结构设计、系统软件以及应用软件设计的共同努力来降低功耗。

4. 带宽障碍。

随着工艺技术的发展，片内的处理能力越来越强。按照目前的发展趋势，现代处理器很快将在片内集成十几甚至几十个高性能处理器核，而芯片进行计算所需要的数据归根结底是来自片外。高性能的多核处理器如不能低延迟高带宽地同外部系统进行数据交互，则会出现“嘴小肚子大”、“茶壶里煮饺子”的情况，整个系统的性能会大大降低。

由于芯片的引脚数目不可能无限增加，现在业界普遍通过高速总线来提高处理器的带宽，不论是访存总线（如 DDR3、FBDIMM 等）、系统总线（如 HyperTransport）还是 IO 总线（如 PCI-Express），其频率都已经达到 GHz 级，有的甚至接近 10GHz。可以预见，不久的将来将出现片外传输频率高于片内运算频率的内外频率“倒挂”现象。即便如此，由于片

内晶体管数目的指数增加,处理器系统结构设计也要面临每个处理器核的平均带宽不断减少的情况。

5. 应用的变化。

在计算机发展的初期,处理器性能的提高主要是为了满足科学和工程计算的需求,非常重视浮点运算能力。随着互联网和媒体技术的迅猛发展,网络服务和移动计算逐渐成为非常重要的计算模式,新的计算模式迫切要求微处理器具有实时响应和处理流式数据类型的能力、支持数据级和线程级并行性、更高的存储和输入/输出带宽、低功耗、低设计复杂性和设计的可伸缩性,同时又要求缩短芯片进入市场的周期。

6. 结构周期律。

在过去 60 年的发展历程中,计算机的体系结构每 20 年左右就出现一个较大突破,已经经历了一个由简单到复杂,由复杂到简单,又由简单到复杂的否定之否定过程。最早期的处理器结构由于工艺技术的限制,不可能做得很复杂;随着工艺技术的发展,出现了流水线技术、动态调度技术、缓存技术、向量机技术,使得处理器结构日趋复杂;精简指令集计算机(RISC)技术的提出使处理器结构得到一次较大的简化;但后来随着工艺技术的进一步发展以及宽发射技术的实现,精简指令集计算机处理器结构变得越来越复杂。近年来精简指令集计算机在微体系结构上没有大的突破,其结构已经非常成熟。现在的精简指令集计算机微处理器普遍能允许几十条到上百条指令乱序执行,如 Alpha 21264 处理器的指令队列最多可以容纳 80 条指令,Power4 为 200 多条。目前,包括超标量精简指令集计算机和显式并行指令计算(EPIC, Explicitly Parallel Instruction Computing,)在内的指令级并行技术使得处理器核变得十分复杂,通过进一步增加处理器核的复杂度来提高性能,空间已经十分有限。同时,结构上通过细分流水线来提高主频的方法很难再延续下去。因此,传统的高主频复杂设计遇到了越来越严重的障碍,需要探索新的结构技术在简化结构设计的前提下充分利用摩尔定律提供的晶体管,以进一步提高处理器的功能和性能。

上述工艺、结构、功耗、应用等几个方面的趋势,使处理器的微体系结构如不进行突破性的变革就无法满足未来的需求,这场变革应该是由复杂到简单的变革,应该面向网络服务和媒体的应用,应该考虑低功耗的要求,应该采用层次的结构以简化物理设计的复杂度。正是由于其较好符合上述趋势,片内多核及多线程技术目前正在受到业界越来越多的关注。

3.2 国外发展现状

目前国外高性能微处理器的研制有两个明显的趋势:一是高性能处理器在技术上和市场上的竞争都很激烈,经过市场的洗牌,处理器的研制越来越集中到少数几家大厂商;二是单处理器性能的继续提高在主频、结构、功耗等方面都碰到了明显的障碍,各微处理器公司都纷纷推出多核结构的微处理器。

九十年代末生产主流通用处理器的厂家很多,如英特尔、AMD(超微)、IBM、惠普(HP)、SUN(升阳)、DEC、SGI 等公司,形成了在通用处理器市场群雄逐鹿的局面。然而,经过近十年的市场洗牌,很多公司退出了通用处理器领域的市场竞争。最早退出通用处理器市场的是 DEC 的 Alpha 处理器。Alpha 处理器无论从结构设计还是物理设计的角度都堪称技术的经典,甚至被很多教科书所采用。但由于 DEC 太专注于技术而忽视市场,在其实现 Alpha 21464 过程中,1998 年被康柏(Compaq)收购了。随后惠普公司又合并了康柏公司,却同时碰到一个棘手的问题,即在扩展基于新的 64 位构架显式并行指令计算(IA64)系统生产线的同时,是否还要保留另外两个精简指令集计算机系列的产品(PA-RISC 和 Alpha)。目

前，惠普公司已经宣布所有 Alpha 处理器开发和服务都将逐步中止，而原先 Alpha 的队伍分别加入英特尔和 AMD；同时，惠普公司在 2004 年发布双核的 PA8800 芯片后也停止了 PA-RISC 的研发。生产 MIPS 系列高性能处理器的 SGI 公司以 2002 年宣布停止 MIPS 18000 的研发为标志，也退出了处理器研发的历史舞台。

在剩下的几个高性能处理器生产厂商 IBM、英特尔、AMD 和 SUN 中，英特尔垄断着 X86 市场，主要面向 PC 和服务器，IBM 的 Power 系列主要占据大型机、服务器等高端市场，SUN 公司的 SPARC 系列则是在颠簸中前进。下面对上述几个公司在多核处理器研发方面的情况做简单介绍。

1. IBM

是最先推出多核微处理器的厂商，其 Power4 已成为多核处理器的代表性产品。IBM 在蓝色基因巨型机中使用自己的双核芯片，奠定了其在多核微处理器研制方面的领先地位，Power 系列处理器主要用于高性能服务器和适度规模并行计算机系统。

2001 年发布的 Power4 片内集成两个类 Power3 处理器核，每个核为 8 路超标量处理器乱序执行，一级指令和数据缓存分别为 64KB 和 32KB，两个处理器核通过交叉开关共享片内 1.5MB 的二级缓存。Power4 采用 180nm 绝缘硅工艺和铜互连技术，片上集成 1.74 亿晶体管。2004 年 IBM 发布了 Power5，Power5 仍集成两个处理器核，每个核支持两个同时多线程（SMT），改造为同时多线程仅增加了 24% 的芯片面积。Power5 片内集成了 1.92MB 的二级缓存、三级缓存的目录以及内存控制器。Power5 采用 130nm 工艺，集成了 2.76 亿晶体管，主频 1.90GHz。2007 年发布的 Power6 采用 65nm 绝缘硅工艺，主频接近 5GHz，采用 DSL^[4] 技术。Power6 主要通过电路设计提高主频，同时把 Power4/5 中共享的二级缓存变成私有的二级缓存，每个处理器核有 4MB，片内增加了一个内存控制器，使得处理器的带宽达到 300GB/s。

2. 英特尔

于 2004 年推出 Montecito，主要面向服务器应用。Montecito 是双核多线程处理器，每个处理器核在 Itanium2 的基础上增加了 2 路阻塞多线程机制。Montecito 采用 90nm 工艺，集成 17.2 亿晶体管，在晶体管数量的竞赛中遥遥领先，其每个处理器核包含 12MB 三级缓存。

2005 年英特尔推出基于 P4 系列的双核处理器 SmithField，将两个 P4 的处理器直接封装在芯片上，两个处理器核共享前端总线。2006 年初，英特尔推出基于 Banias 的双核处理器 Yonah，两个处理器核共享二级缓存和系统接口，可以部分消除 SmithField 中前端总线竞争冲突问题。2006 年英特尔推出了基于 Core 构架的处理器 Conroe（酷睿 2）。Core 构架的主要改进之处包括：增加流水线宽度（由原来的处理三条 x86 指令变为处理四条）、增强 SSE^[5] 功能（由 64 位通路增加到 128 位）、支持 128 位的读指令、动态调节共享缓存的分配等。英特尔的测试数据表明 Conroe 比上一代处理器在性能上提高了 40%，而功耗降低了 40%。2007 年英特尔推出了四核处理器 Core 2 Quad（称为 Kentsfield），实际上是把两个 Conroe 处理器封装在一个芯片上（也称为多芯片封装），含有 8MB 的二级缓存。后续的 Yorkfield 处理器将采用 45nm 工艺，共享 12MB 的二级缓存，支持 1333MHz 的 FSB 以及 DDR3。

英特尔的下一代八核处理器为 Penryn 和 Nehalem。Nehalem 的每个处理器核支持两路同时多线程，整个处理器同时支持 16 个线程，八个处理器核共享片上缓存，片上集成内存控制器，其技术路线开始向 Alpha 和 AMD 靠拢。

3. AMD

AMD 和英特尔一直是全球处理器市场上的两大宿敌。AMD 发展 64 位 X86 和双核处理器的步伐走在了英特尔前面，2004 年推出了双核 Opteron。双核 Opteron 片内集成 2 个 K7 处理器核，每个处理器核有独立的二级缓存，通过交叉开关（Crossbar）实现处理器核和系统请求接口互连，片内集成内存控制器和输入/输出控制器（三个 HyperTransport 接口）。双核 Opteron 采用 90nm 工艺，晶体管数量在 2 亿个以上，功耗小于 95W。2007 年 AMD 推出四核处理器 Barcelona，结构上增加了一个四核共享的 2MB 的三级缓存，采用 65nm 工艺，晶体管数量为 4.63 亿，主频为 2.2-2.8GHz，功耗为 95W。

AMD 的八核处理器命名为 Sandtiger，采用 45nm 工艺，共享 6MB 三级缓存，支持直接互连构架，支持 HT3.0 和 DDR3，片上集成四个 HyperTransport 接口，预计 2009 年推出。

4. SUN

公司于 2004 年上半年发布了它的第一款双核微处理器 UltraSPARC IV，并在下半年推出了 UltraSPARC IV+。UltraSPARC IV 采用芯片多线程（CMT，chip multithreading）技术，片上集成了两个 UltraSPARC III 的内核、二级缓存的标签（tag）体和内存控制单元，外部缓存为 16MB，每个内核独享 8MB，采用 130nm 工艺，主频为 1.2GHz。

2004 年 Sun 公司公布了 Niagara，也称为 UltraSPARC T1，包括八个简单的单发射六级流水的处理器核，每个核支持四个线程，共支持 32 个线程。2006 年，Sun 公司推出了 Niagara-2，仍然为八个 Sparc 处理器核，但每个处理器核增加了浮点部件，每个核支持八个线程，共享 4MB 的二级缓存，片上集成四个双通道的 FBDIMM 内存控制器。

4 纳米级工艺对系统结构的新挑战

多核处理器的发展是工艺推动的结果，面临理论、系统以及物理实现多方面的问题，这些问题既是必须面临的挑战，也是难得的发展机遇。

4.1 多核处理器研究的最新进展

除了上述各大公司推出的多处理器芯片产品以外，国际上多处理器芯片的研究项目也非常多，比较典型的包括以下几个。

1. **IBM 开发中的千万亿次（PetaFLOPS）计算机蓝色基因（Blue Gene）中的核心处理芯片**是探索新型高性能微处理器体系结构的重要一步。该芯片中含 32 个简单的 PowerPC 440 处理器及内嵌的 DRAM，峰值性能可达 32GFLOPS。将 DRAM 集成在片内使处理器到存储器的访问延迟大为缩短，访问带宽大大提高。为了通过精简设计提高运算速度和降低功耗，每个处理器能执行的指令的数量都大大地减少了，但每个处理器仍可支持八个同时执行的线程以提高处理器利用率。片上还集成了六个 2GB/s 带宽的通道以实现与系统中其他同类芯片的通信。另外，为了提高处理器的可靠性，该处理器在硬件上采用了大量的冗余设计，在软件上增加了分布式控制和恢复的机制。
2. **MIT 研究的可重构 RAW^[6] 处理芯片**在单个芯片上将几百个非常简单的处理单元用可重构逻辑连接起来，实现高度并行的体系结构。RAW 允许编译器或其他相当的软件工具重新构建硬件体系结构的低层细节，对每个要加速的应用实现最佳的资源分配。这种结构设计简单，单元内部和单元之间的互连线短，能充分地支持流水线并行性，特别适于未来线宽变窄的半导体制造工艺。RAW 的一个重要特点在于指令执行前的读寄

寄存器部分除了常规从寄存器中取出的操作数和从功能部件刚算出的操作数外，还有通过通信网络从别的处理器那里送过来的操作数。在灵活的互连网络的配合下，这个机制巧妙地把多个处理器的功能部件耦合在一起，构成功能复杂、动态重构的并行处理系统。RAW 的研究者认为，开始时 RAW 体系结构还只适于流式的信号处理计算，但 RAW 方法未来将发展成为普适的解决方案。

3. **Cell 高性能处理芯片。**在 2005 年国际固态电路会议 (ISSCC) 上，IBM、索尼 (Sony) 和东芝 (Toshiba) 首次公开介绍了 Cell 高性能处理芯片。Cell 的一个主要用途是索尼的第三代 Playstation 游戏机 PS3，但 IBM 同时也宣称 Cell 将是片上超级计算机。Cell 可在 4GHz 频率下工作，其宣称的峰值浮点运算速度为 256GFLOPS (单精度)。Cell 由一个相对比较简单支持同时多线程并行的双发射 64 位 PowerPC 内核 (称为 PPE) 和八个 SIMD^[7] 型向量协处理器 (称为 SPE) 构成。片内有一个高带宽的环状高速总线 (EIB) 把 PowerPC 内核、向量协处理器和存储器总线 (RAMBUS) 内存接口控制器 (MIC) 以及 Flex 输入/输出外部总线接口控制器 (BIC) 连接起来。PowerPC 内核主要负责控制并执行操作系统，向量协处理器完成主要的计算任务。向量协处理器的 SIMD 执行部件是 128 位宽的，可在一个时钟周期里完成四个 32 位的定点或浮点乘法运算。向量协处理器内置了 256KB 的 SRAM^[8] 作为局部存储器 (它的编址独立于片外的 DRAM^[9])。
4. **IBM 和德克萨斯州立大学联合开发中的 TRIPS 芯片**是设计具有每秒万亿次运算能力芯片的第一个尝试。TRIPS 结构采用粗粒度的处理器内核，以便在有较高指令级并行性的单线程应用上实现更高的性能；并在同一芯片上重复设置许多这样的内核。片上还集成了存储部件和通信部件，并允许软件调度程序对它们灵活配置，以获得最佳性能。TRIPS 的目标是在 35 纳米的工艺条件下，达到 5TFLOPS 的峰值浮点运算性能。TRIPS 同时兼顾桌面应用和服务器应用，希望到 2010 年，能把桌面 PC、高性能计算、数字信号处理、服务器应用等统一到同一类型的微处理器芯片上。
5. **英特尔开发万亿级研究芯片。**英特尔在其 2006 年开发者论坛中首次介绍了英特尔的“万亿级” (TeraFLOPS) 研究原型芯片。“万亿级”研究原型芯片是世界上第一个达到每秒 1 万亿次浮点运算的处理器，包括 80 个处理器核，频率 3.1 GHz。该芯片综合了英特尔公司近来在硅光子学方面的技术突破，达到了万亿次的性能、每秒万亿字节的访存带宽以及每秒万亿比特的输入/输出速度。该芯片主要结构特点是：(1) 包括 80 个瓦片 (tile)，排成 8×10 的行列结构，每一个瓦片都包括一个带简单指令设置的小型核心用于处理浮点数据；(2) 有一个 20MB 的 SRAM 内存芯片被重叠绑定在处理器核上，这种结构使数百个芯片互连变得比较容易，并且可在内存和核心之间提供超过每秒万亿字节的带宽；(3) 采用混合硅激光芯片，该技术可以形成每秒万亿比特的光学通路。

4.2 多核处理器面临的基本理论问题

目前的计算机结构以图灵机理论为基础，属于冯·诺依曼体系结构。本质上，图灵机理论和冯·诺依曼体系结构是一维串行的，而多核处理器则属于分布式离散的并行结构，如何解决二者的不匹配成为计算机理论和技术的一个根本问题。

1. **串行的图灵机模型和物理上分布实现的多核处理器的匹配问题。**图灵机模型意味着串行的编程模型。串行程序很难利用物理上分布实现的多个处理器核获得性能加速。同时，目前并行编程模型并没有获得很好的推广，仅仅局限在科学计算等有限的领域。研究者应该寻求合适的机制来实现串行的图灵机模型和物理上分布实现的多核处理器

的匹配问题或缩小二者之间的差距，解决“并行程序编程困难，串行程序加速小”的问题。目前主要的研究包括以下几个方面：

- (1) 引入新的能够更好地表示并行性的编程模型。由于新的编程模型支持编程者明确表示程序的并行性，因此可以提升性能，比如 Cell 处理器提供不同的编程模型用于支持不同的应用。其难点在于如何有效推广该编程模型以及如何解决兼容性问题；
 - (2) 提供更好的硬件支持以减少并行编程的复杂性。为了避免并行程序中用复杂的锁机制实现对临界资源的同步及互斥操作，TCC (Transactional Coherence and Consistency, 事务处理一致性协调机制) 机制以事务 (Transaction) 而不是单个访存操作为单位考虑数据一致性问题，简化了并行编程的复杂度；
 - (3) 利用多核加速串行程序。利用软件或硬件自动地从串行程序中派生出能够在多核处理器上并行执行的代码或线程，主要有并行编译器、推测多线程以及基于线程的预取机制等方法。
2. **冯·诺依曼体系结构的一维地址空间和多核处理器的多维访存层次的匹配问题。**冯·诺依曼体系结构采用了一维地址空间，而多核处理器中由于不均匀的数据访问延迟和同一数据在多个处理器核上的不同拷贝产生了数据一致性问题。解决这个问题的研究分为两大类：
- (1) 引入新的访存层次。新的访存层次采用一维分布式实现方式。典型的例子是增加分布式统一编址的寄存器网络，全局统一编址以及只有一个备份的机制避免了复杂的数据一致性问题，而且寄存器还能提供更快的访问速度；TRIPS 和 RAW 都实现了类似的寄存器网络。新的访存层次也可以是私有的形式，每个处理器核都有自己私有的访存空间；Cell 处理器为每个向量协处理器核设置了私有的数据缓冲区；
 - (2) 研制新的缓存一致性协议。其重要趋势是放松正确性和性能的关系。比如推测缓存协议在数据一致性未得到确认之前就推测执行相关指令，从而减少长延迟访存操作对流水线的影响，令牌一致性 (Token Coherence) 技术和前面提到的 TCC 也采用了类似的思想。
3. **程序的多样性和单一的体系结构的匹配问题。**未来的应用展现出多样性的特点。一方面，处理器的评估不仅仅局限于性能，也包括可靠性，安全性等其他指标；另一方面，即便考虑仅仅追求性能的提高，不同的应用程序也蕴含了不同层次的并行性。应用的多样性驱使未来的处理器具有可配置、灵活的体系结构。TRIPS 在这方面作了富有成效的探索，其处理器核和片上存储系统均有可配置的能力，从而使得 TRIPS 能够同时挖掘指令级并行性、数据级并行性及指令级并行性。

4.3 多核处理器面临的系统问题

除了上述理论问题之外，多核处理器也面临一些系统问题，包括处理器的访存瓶颈，可靠性问题，功耗问题等。

1. **处理器的访存瓶颈 (Memory Wall)。**随着片内处理能力的不断提高，传统的计算与存储分离设计使得访存带宽和速度成为未来提高处理器性能的主要瓶颈。根据片上并行处理技术的发展趋势，不久就会出现当处理器发射数百甚至上千条指令时只能取一个值到片上存储器的情况。因此，要求将处理器和存储器做在同一块芯片上，把应用执

行期间的工作集尽可能放到片上。预取技术可以缓解访存瓶颈，传统的硬件或软件预取适合于具有规则访存行为的应用程序。在多线程环境下可以利用空闲的多线程硬件资源通过动态执行地址预测代码进行预取，不仅可以加速具有规则访存行为的应用程序，而且能够有效地加速访问模式不规则的应用。多线程处理器还可以通过线程切换来容忍访存延迟。单个芯片上的输入/输出引脚数目由于物理技术的限制不能同晶体管数目同比例增长，使得计算能力的增长与片外带宽的增长不匹配。从 180 纳米到 35 纳米，晶体管数与信号引脚数的比率增加了 45 倍。有限的片外带宽将成为影响性能和扩展性的瓶颈，并限制了单个芯片上处理器数目的有效增加（尽管有更多的晶体管资源可用）。因此，在结构设计时，我们需要寻找在一定带宽上可容纳的处理器核数、处理器核的复杂度、不同应用的存储模式、以及片上存储容量之间的平衡关系。

2. **可靠性问题。**随着摩尔定律的延续，芯片特征尺寸进一步按比例缩小，在单芯片上集成数十亿晶体管已成为可能。与此同时，由制造过程中工艺参数的涨落或内部原子级效应引起的器件参数离散性不断增加；越来越薄的栅氧化层导致因隧道穿越效应而影响可靠性的机率不断增加；越来越多的片内存储器（单个存储单元的尺寸越来越小）给单粒子效应等因素导致的软失效的机会不断增加。随着缺陷密度增加，芯片的成品率不断下降、芯片的失效速率不断攀升，使芯片的高可靠性设计成为延长硅集成电路摩尔定律最迫切需要解决的问题之一。传统的考虑故障容忍的容错方法成本较高，且其有效性受到失效速率上升的严重影响。必须研究使电路和系统从故障中自动恢复的新原理，从缺陷容忍、故障容忍和差错容忍等方面研究支持芯片高可靠设计的新结构、新方法，从而提高芯片成品率，降低成本，构造稳定可靠、性能可预测的系统。多核处理器上可以采用线程级的容错方法，即通过硬件或软件自动产生冗余线程以确保程序运行的正确性，如 Slipstream Processor 以及 DIVA 处理器均采用类似思想。
3. **功耗问题。**随着摩尔定律的进一步延续，芯片产生的功耗与封装及系统允许的功耗之间会产生巨大的缺口。功耗问题已经成为微处理芯片设计第一位的限制因素。就像提高处理器性能需要从算法、系统、结构、电路、工艺等多个层次来考虑一样，降低功耗也需要从上述多个层次不同角度进行系统研究。例如，可以通过设计新的算法减少程序执行过程中运算操作的次数以减少能量消耗；可以采用异构处理器结构，使得简单处理器核运行简单任务，复杂处理器核运行复杂任务，从而取得合理的能量消耗；可以由编译优化减少访存次数来减少功耗；可以通过操作系统关闭当前不用的功能部件及处理器核，或者动态调整处理器核的时钟频率及电压达到控制功耗的目的；可以重新设计总线的编码方式，减少同时跳变的信号个数来降低功耗；可以在逻辑和电路层次上通过门控时钟、门控电源、双阈值电压、动态电压变换、半频率时钟、异步逻辑等方法降低功耗；可以通过工艺进步减少耦合电容以及降低漏电来降低功耗等。

随着晶体管特征尺寸的缩小和芯片规模的扩大，不仅片上互连线的延迟远远超过了晶体管的开关延迟，而且其能耗已经超过了晶体管电容充放电所消耗的能量。斯坦福大学的威廉姆·达利（William Dally）教授指出，在 130nm 和 50nm 工艺下，在片上 10mm 长度传输 32 位数据所消耗的能量与一次 32 位算术逻辑单元（ALU）操作、一次寄存器读操作和一次 8KB 存储器读操作所消耗的能量，对应比例分别为 20:1:10:2 和 56:1:28:5.6。可见互连线设计是纳米级芯片设计的难点和重点。低功耗互联网络设计也是多核处理器的重要研究方向之一。

4.4 多核处理器面临的物理实现问题

1. **纳米级集成电路的物理设计和分析技术。**纳米工艺对芯片物理设计在各个方面都提出

了更高的要求，其中所涉及的关键技术包括以下几个方面：

- (1) 全芯片规模静（动）态信号完整性分析。深亚微米芯片中连线尺寸缩小，相互间耦合电容所占比重加大，连线间的信号串扰日趋严重。需要通过预防（如限制最大线长）、分析及修复等手段防止线间串扰对正确性或性能的影响，并在信号完整性分析中避免由于过于保守而牺牲性能；
 - (2) 全芯片功耗动态分析方法。纳米级集成电路芯片内部功耗和压降的动态变化日趋剧烈，除了传统的静态分析以外，需要对芯片的功耗进行动态分析。全芯片的动态功耗分析不仅要考虑芯片内部电路、连线的功耗变化，而且要考虑封装所带来的功耗和压降。全芯片功耗动态分析要求用于功耗分析的功耗测试向量能够准确（不是过于保守地）反映芯片工作过程中各个部分的功耗及压降变化；需要运算规模、运算时间与计算精度间的有效平衡；
 - (3) 芯片整体静（动）态热分析。纳米级集成电路尤其是高性能处理器的集成度较高、运行速度快，导致其发热量很大，对芯片的性能和可靠性都有极大的影响。为了对芯片运行过程中的各个部分热量释放及温度变化有一个较为准确的分析，需要建立完善的单元及器件热源产生和传导模型；建立有效的热分析手段分析和捕捉芯片工作过程中内部热量变化和热点；建立反馈途径，根据分析结果修改设计或对封装提出要求；
 - (4) 多工作环境低偏差时钟树的设计。纳米级芯片上的性能参数（如介电常数、掺杂浓度等）的漂移变化会导致时钟树产生很大偏差（Clock Skew），需要结合不同工作环境下的晶体管性能参数变化，对时钟树的结构进行优化调整，保证在各种工作环境下达到时钟偏差的最小化和均衡化，保证芯片性能的可靠和稳定。
2. **纳米级集成电路的全定制设计技术。**在定制设计方面，纳米级工艺的晶体管尺寸进一步缩小，栅氧厚度只有几个原子的高度（10Å左右），导致阈值电压变小（0.1v~0.2v），使得下列问题更加突出：
- (1) 晶体管呈现出非常强的非线性特性，SPICE^[10]模型更加复杂，需要在仿真时对晶体管的特性有更精确的把握；
 - (2) 漏电流随晶体管特征尺寸的缩小以指数速度增大。源漏的漏电流功耗占总功耗的比重已经非常大，同时对可靠性提出了更高的要求。栅极的漏电流已不可忽略，在功能仿真中要考虑到由于栅漏电导致的失效；
 - (3) 对干扰更敏感，抑制噪声能力显著下降。这不仅对仿真，同时也对版图设计提出了更高的要求，需要通过在敏感信号线周围采用电源地屏蔽、拉开间距等方法减小耦合；
 - (4) 深亚微米工艺掺杂浓度更大，导致电流增大，自热现象和电迁移现象都很严重。当电流经过金属线时，电子会撞击金属原子并产生热量。如果在一段时间内有大量的电子撞击金属原子，金属原子会在电流方向上发生位移，移动的结果使金属线变细甚至断开，或者和另一条相邻的金属线发生短路。这要求建立精准的电迁移模型，精确分析电流密度，做更好的预防处理；
 - (5) 闩锁效应（Latch-up）也随着电流增大变得更加严重，衬底噪声变得不可忽略。全定制中的电源/地的分布需要更多地考虑衬底/阱的接触电阻，减小 β 值。

3. **高速信号传输技术。**高速信号传输的关键技术包括以下几个方面：（1）纳米工艺对高速混合信号电路设计的挑战；（2）高性能的射频电路和混合信号电路设计：如低抖动锁相环/延迟锁相环（Low Jitter PLL/DLL Phase Lock Loop/Delay Lock Loop），展频时钟（SSCG ,Spread Spectrum Clock Generator），高采样率的模数转换（ADC），复接器和解复接器（MUX and DEMUX）等的设计；（3）自适应的阻抗匹配和信道均衡设计，来满足日益增加的信号传输速度对信号完整性的要求；（4）可靠的抗静电保护系统（ESD, Electrostatic Discharge）系统同高速输入/输出之间的设计折中；（5）高性能前提下的低功耗设计技术尤其是混合信号电路中的动态功耗管理；（6）在变化的工作环境和偏差越来越大的工艺条件下，保证芯片的可靠性和成品率。

5 国内发展现状及发展建议

5.1 国内发展现状

中国经过多年的努力，特别是“十五”科技计划的实施，集成电路设计与制造产业都得到了快速发展。在集成电路设计方面，2005年销售收入首次突破100亿，达到124.3亿元，同比增幅为52.5%；不仅拥有相当数量的无制造集成电路产品企业（Fabless），还涌现出一批产值超亿元的骨干企业群体；在微处理器、数字信号处理器和片上系统（SOC）等关键战略产品领域取得了重大的进步；出现了如电信智能卡芯片、第二代居民身份证专用芯片等处于世界领先水平的量大面广产品。在集成电路制造方面，出现了以中芯国际、宏力、华虹等为代表的集成电路代工厂，其工艺水平达到了0.18 μm 或0.13 μm ，有的甚至达到90nm。

随着集成电路设计和制造技术的发展，中国微处理器、数字信号处理器的研发能力也取得了重大进步，具备了较好的研发基础。加快自主知识产权微处理器的研究开发和产业化，突破微处理器这一制约中国信息产业良性自主发展的瓶颈，获得信息产业发展的主动权，已经成为党和国家及社会各界的广泛共识。经过长时间的积累和科技人员的不懈努力，中国的微处理器研发已经从主要面向军事应用开始向民用转移，从较低性能微处理器向高性能微处理器迈进，从模仿、仿制向自主开发迈进，开始出现重大的、战略性的转变。

与国外微处理器设计厂家不断合并相对应的是国内微处理器设计能力在“十五”期间的快速增长。在“十五”期间国内微处理器设计主要包括以下特点：（1）通用处理器和嵌入式处理器的研发都在蓬勃发展；（2）产业化开始起步，但形成规模还需要时间；（3）研发还主要集中在单处理器核的设计，部分研究单位在单处理器的研发方面已经达到很高的水平，在“十一五”期间应展开多核处理器的研发。

在国家“八六三计划”和中科院知识创新工程资助下，中科院计算所从事龙芯系列高性能通用处理器的研制，先后完成龙芯1号、龙芯2号、龙芯2号增强型处理器（简称龙芯2E）的研制。其中龙芯2E兼容64位的MIPS指令系统，采用四发射的动态超标量超流水线结构，实现乱序执行技术（如转移猜测、寄存器重命名、动态调度等），以及动态存储访问机制（如非阻塞的缓存访问、取数猜测等）。龙芯2E片内含64KB/64KB一级指令/数据缓存以及512KB二级缓存，片内集成了DDR^[11]内存控制器。龙芯2E采用意法半导体90nm工艺设计，主频达到1GHz，具有低功耗、高安全性特点，SPEC CPU2000的定点/浮点实测分值均超过500分，与中低档的英特尔P4处理器相当，是目前除了美国之外世界上性能最高的通用微处理器。在中法两国政府信息领域战略合作协议的指导下，意法半导体公司购买了龙芯2E的生产和销售授权，取得了很好的经济效益，这是中国在计算机领域的核心技术首次对外授权。基于龙芯2E的低成本电脑、龙芯笔记本等产品已经完成万套规模的试用，正在

进入规模推广阶段；比龙芯 2E 的集成度更高、成本/功耗更低的龙芯 2F 已经研制出来，并正在进行百万片级的量产。龙芯 2F 的平均性能比龙芯 2E 高 20% 以上，输入/输出带宽比龙芯 2E 高出一倍，但功耗降低了 40%。基于龙芯处理器的高可靠计算机模块也通过了可靠性实验，在国家战略安全领域取得了重要应用。

在十五期间龙芯 2 号成果的基础上，中科院计算所正在进行多核的龙芯 3 号的研究。龙芯 3 号将采用 65nm~45nm 工艺设计，实现四核和 16 核两个主要产品芯片，每个处理器核采用四发射结构，主频在 1GHz 以上。计划 2008 年完成四核龙芯 3 号的设计；2009 年完成 16 核龙芯 3 号的设计。此外，中科院计算所还在“九七三项目”的支持下，进行片内万亿次领域通用处理器芯片的研发。目前，中科院计算所 65nm 的测试芯片已经流片成功。

此外，北京大学、同济大学等单位也已研制成功嵌入式微处理器以及有关的片上系统。

5.2 发展建议

1. **抓住机遇实现跨越发展。**随着集成电路制造技术进入纳米级，目前国际上高性能通用微处理器的发展正面临技术转型期。性能功耗比继性能价格比之后正在成为计算机的主要设计指标，主频至上的计算机处理器设计技术正在终结，互联网的普及正在改变计算机的应用模式，电子设计自动化（EDA，Electronic Design Automatic）工具不断完善和成熟以及集成电路代工厂蓬勃兴起。上述技术转型一方面降低了我们进行高性能处理器设计的门槛，另一方面加大了国外处理器设计厂商进一步前进的难度，为中国在未来几年发挥后发优势，另辟蹊径，通过跨越创新实现突破提供了机遇。我们应该紧紧抓住上述机遇，加大研发和产业化力度，实现跨越发展。
2. **重视新材料和新工艺研发。**材料和工艺技术是集成电路的基础，虽然传统的硅技术还有一定的发展空间，但进一步发展的难度越来越大，而且迟早会遭遇物理极限。传统的 CMOS 工艺受到制造成本的制约，还能走多远是个未知数。很多工艺厂家已经明确宣布其工艺发展的极限，例如德州仪器（TI）已宣布只做到 45 纳米工艺，IBM 宣布不考虑 22 纳米以后的 CMOS 工艺生产线。

中国“十五”期间在以硅为主的集成电路工艺方面进行了大量的投资，在中长期科技发展规划中也设立了专门的重大专项进行集成电路工艺技术的开发。如前所述，国际上主要的半导体集成电路生产工艺厂商都在大力开展基于硅的新工艺研发以及新材料的研发，这些技术和材料有的甚至进入了产品阶段。一旦国际上出现新的材料和工艺替代目前的硅技术，我们辛辛苦苦建立起来的集成电路生产和设计能力就会过时，使我们永远处在 IT 产业链的末端。因此，我们应该未雨绸缪，加紧新材料和新工艺的研发及其推广应用，争取在未来的竞争中占据有利地位。

所谓纳米技术有两条路，目前的 CMOS 技术走的是从顶向下的路，这条路越来越难走了。另一条路是自底向上的路，量子器件或其他纳米器件不一定需要光刻等技术，也许制造成本会低于纳米级的 CMOS 工艺。我们应同时在两条路上努力，不能死心眼，拼投资强度，要以智慧取胜，加强基础研究。目前中国的纳米技术研究，主要偏重在纳米材料，实际上纳米技术最有市场前景的是纳电子技术，国家应加强对纳米信息技术研究的支持。

3. **重视结合部工作。**从技术和实现的角度看，国内处理器设计与国外同行相比，最大的差距不在于单项技术（包括微处理器体系结构、电路设计、工艺、验证、系统软件），而在于这些技术之间的结合部，在于对这些技术能够融会贯通的人才的缺乏。如国内

部分研究单位在处理器的微体系结构研究方面有很强的实力，但真正把这些技术在软件和操作系统、编译系统等系统软件以及在物理上和硅片上晶体管、连线的行为结合在一起考虑则少有研究，技术整合实现上能力不高。又如，中国在“十五”期间工艺技术取得了长足的进步，集成电路代工厂蓬勃兴起，但普遍缺乏芯片设计需要的包括各种单元库、高速输出输入、锁相环等知识产权核数据库（IP）环境，阻碍了国内的集成电路设计的成长。

从产业化和应用的角度看，国内与国外同行相比的最大差距，在于产业环境的不完善。处理器芯片虽然是计算机系统的核心部件，但决不是计算机系统的全部。计算机系统还包括有关的套片、各种接口芯片、系统软件以及相应的应用等，即使是像英特尔和IBM这样的企业，也没有能力包揽一个系统所有软硬件的设计。越是核心技术，产业链越长，产业环境越复杂。国内处理器芯片研发的一个重大障碍是缺少像微软这样的系统和应用软件厂商。因此政府和产业界应该有一个全面长远的产业整体发展战略，不能只集中于个别环节，而要注意整个产业链的健康持续发展和链条上各个环节的“无缝”联接与整合。

4. **不要妄自尊大，也不要妄自菲薄。**克服浮躁情绪，是中国IT界的首要问题。现在没有人否认微处理器是一个关系到巨大政治经济利益的高科技产品，也没有人不希望尽快改变信息产业“无芯”的尴尬局面，甚至希望“中国芯”一步跨越到世界先进水平；然而，被忽略的是：发展微处理器必须有一个长期的发展战略，必须打持久战。在战略上，我们要自尊自信，相信中华民族的智慧可以做出高性能的微处理器；但在战术上，我们不能自高自大，要谦虚谨慎，戒骄戒躁，踏踏实实，认认真真。如果浮躁主导了决策，会给科技和产业的发展带来危害，我们的差距还会加大。

同时，我们也要充分认识到，自主创新的信心不足是制约我们实现创新跨越的最大障碍。由于我们在计算机系统设计上长期落后，造成了“自主研发的处理器芯片以及计算机系统肯定不如国外产品”的思维定势。“自主性”成为不可靠、性能不好的代名词。我们自主研发的龙芯处理器芯片在国内推广困难，却能授权给欧洲最大的半导体厂商——意法半导体进行生产和销售；基于龙芯处理器的计算机系统在国内只在政府支持下进行了万套规模的试点应用，却从南美和欧洲得到百万套规模的定单，形成所谓“墙内开花墙外香”的怪现象。这种状态的改变需要业界、政府和公众的共同长期努力。

6 结束语

随着集成电路工艺技术进入纳米级，高性能微处理器的发展正处在历史的转折点上。主频至上的处理器设计理念正在终结，性能功耗比继性能价格比之后正在成为计算机的主要设计指标，互联网的普及正改变着计算机的应用模式。在快速发展的计算机领域，很少有人能够预测未来5年~10年的发展。不过有两点是可以肯定的：第一、未来5年~10年必将迎来高性能处理器的一个激动人心的发展时期；第二、技术转型期也是后来者的重要机遇期，中国应该紧紧抓住机遇，发挥后发优势，实现跨越发展。

发展具有自主知识产权的工艺技术和微处理器，建立自主可控的计算机工业是国家的需要、时代的需要，是产业发展的需要，并且我们已经初步具备了相应的条件，再没有理由怀疑我们能够做成这件事。但是也应该清醒地看到发展自主可控的计算机工业绝非易事，需要长时间的积累。无所作为的怀疑论和盲目乐观的速胜论都是错误的。“在战略上藐视它，在战术上重视它”是中国发展自主知识产权微处理器和计算机工业时应遵循的一个基本原则。

注释

【1】在应变硅中，晶格里的原子距离比正常晶格大，使原子密度降低，电子通过时的阻力会减少很多。

【2】鳍形场效应管

【3】即一个 4 扇出反相器的延时。

【4】Dual-Stress Line，一种通过在 CMOS 沟道加上不同的应力来提高电子或空穴迁移率的技术。

【5】Streaming SIMD Extensions，是英特尔在 AMD 的 3D Now!发布一年之后，在其计算机芯片 Pentium III 中引入的指令集，是 MMX 的超集。

【6】由 MIT 的 Agarwal 教授领导的微结构项目，目的是开发一种完全分布式模块化的微处理器设计方案。

【7】Single Instruction Multiple Data，单指令多数据流

【8】Static Random Access Memory，静态随机存储器

【9】Dynamic Random Access Memory 动态随机存取存储器

【10】一个商用的电路行为仿真分析软件

【11】Double Date Rate，上下行双数据率

参考文献

- [1] David Patterson, John Hennessy. *Computer Architecture: A Quantitative Approach* (the third edition). Elsevier Science Pte Ltd. 2003
- [2] International Technology Roadmap for Semiconductors, <http://public.itrs.net/>, 2007
- [3] Chau et al. Advanced CMOS Transistors in the Nanotechnology Era for High-Performance, Low-Power Logic Applications. *ICSICT*. 2004
- [4] T. C. Chen, Where Si-CMOS is going: Trendy Hype vs Real Technology. *ISSCC* 2006
- [5] Joel Tandler, Steve Dodson, Steve Fields et al. Power4 system microarchitecture. *IBM Technical White Paper*, October 2001
- [6] Doug Burger, James R. Goodman. Billion-transistor architectures: there and back again. *Computer*, Volume 37, Issue 3, Page (s) :22-28. Mar 2004
- [7] K. Sankaralingam et al. Exploiting ILP, TLP, and DLP with the Polymorphous TRIPS Architecture. *In the 30th ISCA*, 2003, pp. 422-433
- [8] Lance Hammond, Vicky Wong, et al. Transactional Memory Coherence and Consistency. *In the 31st ISCA*, June, 2004
- [9] M. Goma, C. Scarbrough, etc. Transient-Fault Recovery for Chip Multiprocessors," *In the 30th ISCA*, 2003
- [10] W. Wulf and S. McKee. Hitting the Memory Wall: Implications of the Obvious. *ACM Computer Architecture News* 23, No. 1, 20-24 (March 1995)
- [11] J. Goodman, D. Burger, and A. Kagi. Memory bandwidth Limitations of Future Microprocessors. *In Intl. Symp. on Computer Architecture*, 1996, pp. 78-89
- [12] Weiwu Hu, Fuxin Zhang, Zu-Song Li. Microarchitecture of the Godson-2 Processor. *J. Comput. Sci.*

作者简介:

胡伟武:	中国科学院计算技术研究所研究员, 所长助理, “龙芯”处理器设计负责人
李国杰:	中国工程院院士, 中国科学院计算技术研究所研究员、所长