

中国科学院研究生院
2007 年招收攻读硕士学位研究生入学统一考试试题
科目名称：计算机原理

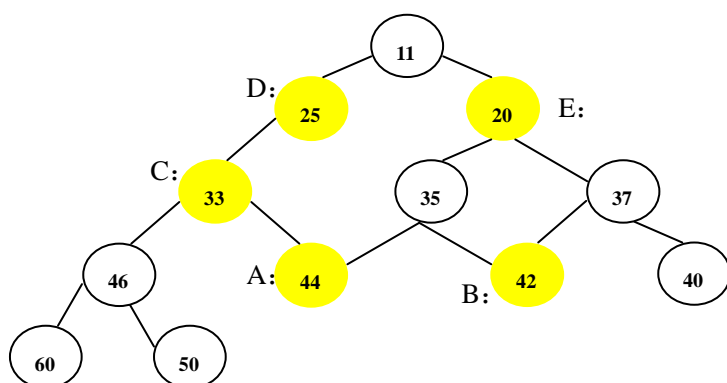
考生须知：

1. 本试卷满分为 150 分，全部考试时间总计 180 分钟。
 2. 所有答案必须写在答题纸上，写在试题纸上或草稿纸上一律无效。
-

数据结构部分（共 75 分）

一、选择题（共 20 分，每题 2 分）：

- 1、假定我们从下图所示的二叉堆中删除了值为 11 的节点，那么值为 50 的节点将出现在图中的哪个指定位置？



- A、A B、B C、C D、D E、E
- 2、有一个 2000 项的表，采用等分区顺序查找的分块查找算法，若每块为 20，平均查找长度为：
 - 3、A 和 B 分别是一棵二叉树中的两个节点，下面说法不正确的是：
- A、A 在 B 的左边，中根遍历时 A 先被访问
- B、A 在 B 的右边，后根遍历时 A 先被访问
- C、A 是 B 的子孙，中根遍历时 A 先被访问
- D、A 是 B 的祖先，先根遍历时 A 先被访问

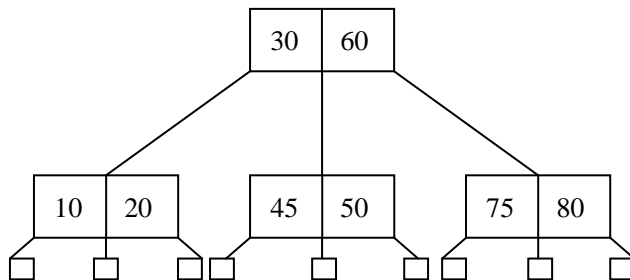
- 4、已知 $A[1..N]$ 是一棵顺序存储的完全二叉树，9 号节点和 11 号节点共同的祖先是：
A、4 B、6 C、2 D、8
- 5、一棵满 3 叉树，按层次遍历的方式存储在一维数组 $A[1..N]$ 中，那么下标为 4 的节点的第 3 个孩子的下标为：
A、10 B、12 C、13 D、8
- 6、在下面的排序算法中哪种是稳定的排序方法：
A、希尔排序 B、快速排序 C、选择排序 D、归并排序
- 7、有 6 个元素按 a、b、c、d、e、f 顺序进栈，下列哪个不是合法的出栈序列
A、bcdaef B、cbdefa C、dcabef D、edcfba
- 8、已知存在广义表 $A=(x, ((a,b),c,d))$ ，在其上进行 $\text{head}(\text{head}(\text{tail}(A)))$ 的结果为
A、(a,b) B、(a) C、(x) D、((a,b))
- 9、假设存在一棵哈夫曼树 T，它具有 m 个页节点，则该树的节点总数为：
A、2m B、m+1 C、2m-1 D 不能唯一确定
- 10、存在一个由 8 个节点组成的图，节点从 0 到 7 编号，图中有 13 条有向边，分别是：0-7 0-1 1-4 1-6 2-3 3-4 4-2 5-2 6-0 6-3 6-5 7-1 7-3。
下面选项中哪个是该图的强连通分量。
A、0-1-4 B、3-5-6 C、0-1-6-7 D、1-4-3

二、简答题（16 分）

(1)、(4 分) 假定我们使用 B-树结构来组织一些位于磁盘上的文件数据。请问为什么 B-树的阶数选择得过大或过小都会使数据查找的性能受到严重影响？

(2) (4 分) 如果我们使用线性探测再 Hash 法为 1000 个元素设计 Hash 表，Hash 函数的类型为 $\text{Hash}(x)=x \bmod D$ 假定我们要求查找成功时平均查找长度不大于 4，不成功时平均查找长度不大于 18.5。那么为了满足上述要求，D 的值最少应为多少。

(3) (4 分) 下图表示的是一个三阶 B-树，请画出在此树中插入值 47 后的结果。



(4)、(4 分)

假设

$$T(n) = \begin{cases} 3 & \text{if } n=1 \\ 2T(n-1) & \text{if } n \neq 1 \end{cases}$$

请写出 $O(T(n))$ ，并证明你的答案。

三、(8 分) 假定在内存中存在着一个由多行组成的文本，每行最多有 255 个字符，并且具有一个唯一标识自己的行号。该文本中的各行以行号的大小从小到大依次存放。要求行号必须连续，且可以从两个方向访问文本中的各行。

试使用链表编写以下函数

- (1) 向文本中插入新行，插入位置由参数给出
- (2) 将文本中相邻的两行物理位置进行交换，注意这里不仅仅是交换两行的内容

四、(8 分) 修改快速排序算法，使它仅输出一个数列中最大的 n 个数，并且这 n 个数不要已排好序。

五、(8 分) 已知存在一个有向图 G ， A 和 B 是 G 中的两个节点，试编写算法求 G 中从 A 到 B 的所有简单路径。假定该有向图使用邻接矩阵的方式存储。

六、(7 分) 已知存在一个数组 $P[N]$ ， $P[i]$ 存放着整数型完全二叉树中第 i 个节点的父亲的节点号，如果 $P[i]=0$ ，则该节点的父节点为根节点。假定二叉树中左孩子的值肯定大于右孩子的值。另外，存在一个数组 $Node[N]$ 存放该二叉树各节点的值。如： $Node[i]$ 存放着第 i 号节点的值。编制一个算法根据 $P[N]$ 和 $Node[N]$ 建构该二叉树的链表表示结构。

七、(8 分) 基于上题所创建的二叉树，设计一个非递归算法查找和等于指定数的所有连通数列（连通序列指该数列对应的节点在树上存在一条连接路径）。假设所提供的数值不能小于等于根节点的数值。例如：输入 12，如果在二叉树中存在值为 1, 3, 8 的节点且这三个节点连通，则将其输出。

计算机组成原理部分（共 75 分）

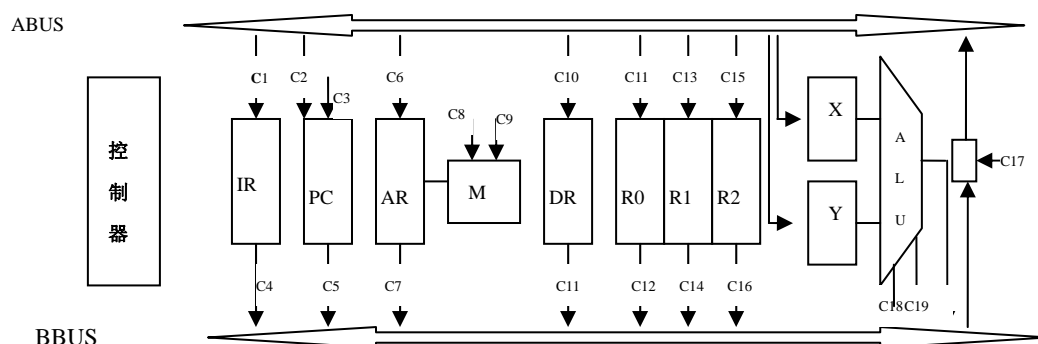
一、填空题：（共 25 分）

（1）（4 分）冯·罗依曼计算机由()、()、()和()组成。冯氏机比起之前的计算装置最大的改进是()。早期的冯氏机以()为中心，现在的计算机是以()为中心，这种变化的主要原因是()。

（2）（3 分）在计算机中，浮点数不能够精确表示，产生这种误差的根本原因是()。在浮点数表示中，阶码一般用移码表示，这主要是因为()。十进制数-5 基于单精度浮点数 IEEE 754 的编码为()（使用十六制表示）。使用单精度浮点数 IEEE754 编码所不能够精确表示的最小正整数为()。

注：单精度浮点数 IEEE754 格式为：符号 1 位、尾数 23 位、阶码 8 位（用移码表示）

（3）（7 分）已知 CPU 数据通路图如下图所示，在线上标有控制信号。



上图中，控制信号 C3 表示_____操作； C8 和 C9 分别表示_____和_____操作；C18 和 C19 分别表示_____和_____操作；

下面列出的是提取指令的操作流程如下所示，请在右侧的空格填写每一步对应的微操作序列

PC->AR _____
M->DR _____
DR->IR _____
PC+1->PC _____

（4）、（3 分）假定一台计算机具有 32 个通用寄存器，该计算机的指令系统可以完成 62 个操作。该指令系统支持直接寻址、寄存器寻址、寄存器间接寻址、基址寻址四种寻址方式。在该指令系统中，有一条整数除法指令，记为 DIVi，该指令的长度为 20 位。对于 DIVi，除数可以存放在任何一个通用寄存器中，商存放在指定的寄存器中，余数也存放在寄存器中。

若采用直接寻址方式，则除数与余数也都放在指定寄存器中，寻址范围为（ ）
若采用寄存器间接寻址方式， 则可供存放余数的寄存器最少有（ ）个
若采用基址寻址方式， 余数存放在指定的寄存器中，地址偏移量的范围为-32~31，可以作为基址寄存器的寄存器有（ ）个。

(5)、(4 分) 假设一台使用页面虚拟存储器的计算机主存容量为 16MB，虚拟存储器容量为 1GB， 页面大小为 4K， 则该计算机的虚拟地址格式为（ ）， 如果每个页表项中包含了有效位、保护位、修改位、使用位共 4 位，那么每个页表项的大小为（ ）位。 页表大小为（ ）位。为了减少页表占用主存的数量，一般机器采用（ ）。假定一个应用程序被分配给 4 个物理页框，页面访问序列为 1 8 1 7 8 2 7 2 1 8 3 8 2 1 3 1 7 1 3 7， 若采用 LRU 替换算法， 缺页率为（ ）。

(6)、(4 分) 一个组相联映像 Cache 由 64 个存储块组成， 每组包含 4 个存储块。 主存包含 4096 个存储块， 每块由 64 个字组成， 访存地址为字节地址。 该 Cache 的地址格式为（ ）、主存地址格式为（ ）主存地址 48AB9H 映射到 Cache 的（ ）组（ ）块。

二、(10 分) 设某计算机有四级中断 A、B、C、D，其硬件排队优先级次序为 A>B>C>D 。下表列出的是执行每级中断处理程序所需的时间：

中断处理程序	所需时间
A	5us
B	15us
C	3us
D	12us

如果我们想以执行中断处理程序的时间作为确定中断优先级的尺度：时间越短优先级越高。
(1) 请指出如何为各级中断处理程序设置屏蔽码？
(2) 如果 A、B、C、D 分别在 6us、8us、10us、0us 时刻发出中断请求，请画出 CPU 执行中断处理程序的序列。
(3) 基于上题，请计算上述四个中断处理程序的平均执行时间。

三、 (10 分)
(1) 简述 DMA 的工作原理。
(2) 在引入 DMA 后对 Cache 的设计产生了什么影响？ 相应问题应如何解决？
(3) 假定磁盘通过使用 DMA 每次向内存传输 8×10^3 B， 磁盘的存取速度为 4×10^6 B/S，假定每次初始化设置 DMA， 需要花费 CPU1000 个时钟周期，CPU 处理中断需要 500 个时钟

周期。CPU 的时钟频率为 $500 \times 10^6 \text{ HZ}$ 。问 CPU 的使用效率是多少？

四、(12 分) 一台计算机有分离的数据和指令 Cache。同时该计算机还采用了页式虚拟存储器技术。这里假定页面和 Cache 块具有相同的大小。已知 Cache 的存取速度为 10ns, 主存的存取速度为 60ns, 磁盘的存取速度为 12ms。该计算机的时钟周期为 10ns。如果指令和数据的提取均命中 Cache, 指令的执行需要 1 个时钟周期。Cache 采用的是直接映射并使用写回策略。在 Cache 中平均 50% 的块是修改过的。对于主存, 同样采用写回策略, 主存中平均 30% 的页面已经被修改。

我们假定指令在 Cache 和主存中的命中率均为 95%, 而数据在 Cache 和主存中的命中率为 90%, 我们还知道一般情况下 35% 的指令存取数据, 求这种情况下的最大 CPI。该题必须写出计算过程, 并对每一步作必要的说明, 否则不给分。

五、(18 分) 假设某计算机的 CPU 采用了一个五段流水线, 该流水线的各段分别为 IF (指令提取)、ID (指令译码及寄存器读取)、Ex (执行)、Ma (访问内存)、WB (将结果写回寄存器)。已知该流水线各功能段的延迟时间均为一个时钟周期。

存在下面一段汇编语言程序

```
LW  R1, 0(R2)      MEM[R2]->R1
LW  R4, 0(R3)      MEM[R3]->R4
MUL R1, R1, R4      R1*R4->R1
ADD R5, R1, R5      R1+R5->R1
ADDI R2, R2, 4      R2+4->R2
ADDIR3, R3, 4       R3+4->R3
BNE  R7, R2, LOOP   if(R7!=R2) Goto Loop
```

注: 斜体字为语句注释

- (1)、试画出上述代码在指定流水线上执行时的时空图
- (2)、在执行这段代码时, 流水线的效率和加速比
- (3)、仍然使用上面的流水线执行另一程序, 该程序中 15% 的指令是条件分支指令, 其中 60% 判断条件成立, 程序执行流程发生转移。另外, 还有 5% 的指令是 JUMP 指令。试回答下列问题:
 - a、如果没有做任何分支预测, 计算由于分支所引起的流水线平均损失时钟周期数。
 - b、如果我们使用静态预测的方法, 假定分支判断条件一直不成立, 计算由于分支所引起的流水线平均损失时钟周期数。
 - c、如果我们使用静态预测的方法, 假定分支判断条件始终成立, 并且即使预测正确也将损失一个时钟周期, 计算由于分支所引起的流水线平均损失时钟周期数。