

2012 April NCKU CSIE PH.D. Qualification Examination
Computer Architecture

1. Some operations on two operands (e.g., subtraction) are not commutative. What are the advantages and disadvantages of the stack, accumulator, and load-store architectures when executing noncommutative operations? (15 points)
2. List all the dependences (output, anti, and true) in the following code fragment. Indicate whether the true dependences are loop carried or not. Show why the loop is not parallel. (20 points)
for (i=2; i<100; i++) {
 a[i] = b[i] + a[i]; /* S1 */
 c[i - 1] = b[i] + d[i]; /* S2 */
 a[i - 1] = 2 * b[i]; /* S3 */
 a[i + 1] = 2 * b[i]; /* S4 */
}
3. The classical approach to improving cache behavior is to reduce miss rates. Please summarize the techniques that can reduce miss rates. (15 points)
4. Explain why the cache miss penalties increase as the processor becomes much more faster than DRAMs? Briefly describe five optimizations to reduce cache miss penalty. (20 points)
5. Describe two major instruction set characteristics that can further divide general purpose register (GPR) instruction set architecture into three classes, based on whether the instruction operands are used explicitly or implicitly. And show the advantages and disadvantages of these three further divided classes. (15 points)
6. Explain the following synchronization primitives: atomic exchange, test-and-set, and fetch-and-increment. Also, explain what is the pair of instructions, load linked (LL) and store conditional (SC) and how this pair of instructions can be used to implement atomic exchange and fetch-and-increment. (15 points)

4/2012 博士班資格考： 機率與統計 Show All Details.

1. (20%)
 - (a) If the RVs X and Y are joint normal with zero mean, find the estimate of parameter y given x . (10%)
 - (b) State the differences among independent, uncorrelated, and orthogonal (10%)

2. (20%) A pollution investigation was made upon a river based on a certain chemical substance measured in milligrams per liter. 15 samples were collected from station 1 and 20 samples were obtained from station 2. The average in station 1 is 3.0 milligrams per liter and a standard deviation being 2.5 milligrams per liter and 1.2 milligrams per liter and 0.5 milligrams per liter for station 2. Find the 95% confidence interval for the difference in the true average substance at the two stations, assuming the observations are all normally distributed.

3. (20%) Let X be a random variable and $Y=g(X)$. Prove that
 - (a) $E[Y] = E[g(x)]$.
 - (b) $f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} + \dots + \frac{f_X(x_n)}{|g'(x_n)|} + \dots$, where $y = g(x_1) = g(x_2) = \dots$

4. (20%)
 - (a) State the conditions of a stochastic process which is called wide-sense stationary.
 - (b) If $x(t)$ is a W.S.S. stochastic process, show that $E\{[x(t+T) - x(t)]^2\} = 2\{R(0) - R(T)\}$.

5. (20%) Let \bar{X} be the sample mean and S^2 be the sample variance of a series of random samples, X_1, X_2, \dots, X_N , of size N from a normally distributed event (Gaussian). (20%)
 - (a) What is the definition of an unbiased estimator of the parameters of the event?
 - (b) Is S^2 the unbiased estimator of the variance? Prove your answer.

Ph.D. Qualify Examination

- This examination is closed books, closed notes
- Please turn off your cell phones
- Write legibly. What can't be read will not be graded
- Good luck!

1. Deterministic Finite Acceptors (DFAs) (5 pts each)

Draw Deterministic Finite Automata to accept the following sets of strings over the alphabet $\{0,1\}$:

- a. All strings whose binary interpretation is divisible by 5.
- b. All strings that contain the substring 0101.

2. Nondeterministic Finite Acceptors (NFAs) (5 pts each)

Draw Non-deterministic Finite Automata with the specified number of states to accept the following sets:

- a. All strings containing exactly 4 "0"s or an even number of "1"s.
(8 states)
- b. All strings such that the third symbol from the right end is a "0".
(4 states)

3. Give regular expressions for the following languages on $\Sigma = \{a, b, c\}$.
(5 pts each)

- I. All strings containing exactly one a
- II. All strings containing no more than three a's

4. Prove that the following languages are not regular (10 pts)

$$L = \{a^n b^l a^k : k \geq n+1\}$$

5. Show that the following grammar is ambiguous. (10 pts)

$$S \rightarrow AB|aaB$$

$$A \rightarrow a|Aa$$

$$B \rightarrow b$$

6. Transform the grammar with productions into Chomsky normal form (10 pts)

$$S \rightarrow abAB$$

$$A \rightarrow bAB|\lambda$$

$$B \rightarrow BAa|A|\lambda$$

7. Construct nondeterministic pushdown automata that accept the following languages on $\Sigma(a, b, c)$ (10 pts)

$$L = \{w : n_a(w) = n_b(w) + 1\}$$

8. Prove that the following languages are not context-free (10 pts)

$$L = \{a^n b^j c^k : k > n, k > j\}$$

9. Show that the class of Pushdown Automata with two stacks are equivalent with the Turing machines (10 pts)

10. Construct Turing machines that will accept the following languages on $\{a, b\}$ (5 pts each)

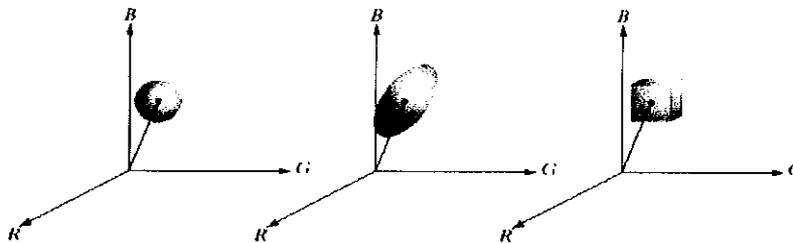
I. $L = \{w : |w| \text{ is even}\}$

II. $L = \{w : |w| \text{ is a multiple of 3}\}$

Digital Image Processing
PhD Qualification Examination
 Department of CSIE, NCKU,

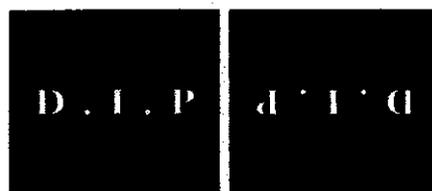
April 2012

1. Give a single intensity transformation function for spreading the intensities of an image so the lowest intensity is 0 and the highest is $L-1$. (20%)
2. What is the deformable model (or snake) algorithm? In comparison with the edge-linking based methods, what are their major similarities and differences? Please also describe at least one case of applying snake algorithm to medical image analysis. (20%)
3. In color segmentation, we can apply the following three diagrams to define the interested color range in the RGB space. Please describe the three approaches with equations and explain how the segmentation is performed. (20%)



a b c
FIGURE 6.43
 Three approaches for enclosing data regions for RGB vector segmentation.

4. Consider the images shown. The image on the right was obtained by (a) multiplying the image on the left by $(-1)^{x+y}$; (b) computing the DFT; (c) taking the complex conjugate of the transform; (d) computing the inverse DFT; and (e) multiplying the real part of the result by $(-1)^{x+y}$. Please explain (each of the five steps) mathematically why the image on the right appears as it does. (20%)



5. Please describe what the Hough transform is. (5%) Please explain why the (r, θ) domain is used instead of the (a, b) domain in Hough transform. (5%) Please describe how to use Hough transform to detect coins which are with three different diameters from an image. (10%)

Discrete-Time Signal Processing 資格考

April 2012

1. (20%) Determine if the systems described by the following input-output equations are (1) linear, (2) stable, and (3) causal.

(a) $y[n] = 4x[n] + 5$

(b) $y[n] = \log(x[n])$

Justify your answer.

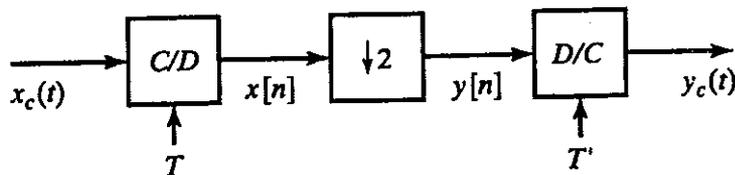
2. (20%) In the following figure, $x[n] = x_c(nT)$ and $y[n] = x[2n]$

- (a) Assume that $x_c(t)$ has a Fourier transform such that $X_c(j\Omega) = 0, |\Omega| > 2\pi(100)$.

What value of T is required so that

$$X(e^{j\omega}) = 0, \quad \frac{\pi}{2} < |\omega| \leq \pi?$$

- (b) How should T' be chosen so that $y_c(t) = x_c(t)$?



3. (20%) Consider a right-sided sequence $x[n]$ with z-transform

$$X(z) = \frac{2z^2 - z}{2z^2 + \frac{3}{2}z + \frac{1}{4}}$$

Determine the inverse z-transform using each of the following methods

4. (20%) Consider a stable linear time-invariant system with input $x[n]$ and output $y[n]$. The input and output satisfy the difference equation.

$$y[n-1] - \frac{5}{2}y[n] + y[n+1] = x[n]$$

- (a) Plot the poles and zeros in the z-plane.
 (b) Find the impulse response $h[n]$.

5. (20%). Suppose that we wish to design a highpass filter satisfying the following specification:

$$-0.04 < |H(e^{j\omega})| < 0.04, \quad 0 \leq \omega \leq 0.2\pi,$$

$$0.995 < |H(e^{j\omega})| < 1.005, \quad 0.3\pi \leq \omega \leq \pi.$$

The filter will be designed using the bilinear transformation and $T=2$ ms with a prototype continuous-time filter. State the specification that should be used to design the prototype continuous-time filter to ensure that the specification for the discrete-time filter are met. The following equations are for your reference :

$$s = \frac{2}{T} \left(\frac{1-z^{-1}}{1+z^{-1}} \right), \quad \Omega = \frac{2}{T} \tan(\omega/2), \quad \omega = 2 \arctan(\Omega T / 2)$$

Qualification Exam (April 2012)

Pattern Recognition

- (15%) Explain the following terms
 - Bayesian Information Criterion
 - Fisher's Linear Discriminant
 - The Bias-Variance Decomposition
- (25%) The broad class of distributions (also called the *exponential family*) of a random vector \mathbf{x} given parameters $\boldsymbol{\eta}$ is written by

$$p(\mathbf{x} | \boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\}.$$

Show that the Gaussian distribution belong to the exponential family and find the corresponding $\boldsymbol{\eta}$, $h(\mathbf{x})$, $\mathbf{u}(\mathbf{x})$ and $g(\boldsymbol{\eta})$. How can you apply the maximum likelihood principle to find the parameters $\boldsymbol{\eta}$ from training data $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and obtain the *sufficient statistics* $\sum_n \mathbf{u}(\mathbf{x}_n)$?

- (15%) What is the objective function of perceptron algorithm? Explain the perceptron convergence theorem.
- (25%) A general linear regression model for expressing multidimensional data $\mathbf{x} = [x_1, \dots, x_D]^T$ is written by

$$y(\mathbf{x}, \mathbf{w}) = \omega_0 + \sum_{j=1}^{M-1} \omega_j \varphi_j(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}),$$

where $\mathbf{w} = [\omega_0, \dots, \omega_{M-1}]^T$ are regression parameters and $\boldsymbol{\varphi} = [\varphi_0, \dots, \varphi_{M-1}]^T$ are basis functions. Given a set of input variables and target variables $\{\mathbf{x}_n, t_n\}_{n=1}^N$, show your procedure of finding the least squares solution and illustrate the relationship between least squares and maximum likelihood. In addition, illustrate the relationship between the regularized least squares and the Bayesian linear regression.

1. (20%) For a set $S \subseteq N$ of size n , determine the number of spanning trees with vertex set S .
2. (20%) Show that the center of a tree is a vertex or an edge.
3. (20%) Solve the Chinese Postman Problem in the k -dimensional cube Q_k under the condition that every edge has weight 1.
4. (20%) Show that if every vertex of a graph has degree at least 2, then G contains a cycle.
5. (20%) Prove or disprove: the independence number of a X, Y -bigraph equals $\max\{|X|, |Y|\}$.

作業系統 資格考題 (100 年度第 2 學期)

1. (20%) Explain the 4 necessary conditions for a deadlock to occur, and suggest a method to break one of the conditions so as to **prevent** deadlocks.
2. (20%) What's the fork() system call? What's the exec() system call?
3. (20%) For the multithreading models,
 - (a) Please explain the 1-to-1 and the many-to-1 models.
 - (b) What are the benefits of the 1-to-1 model, compared to the many-to-1 model?
4. (15%) Given a 10,000-RPM disk with 80-MB/second bandwidth and 10-ms average seek time, please calculate the average time to read a 40-KB block from this disk.
5. (15%) Describe how the Shortest Job First (SJF) scheduling algorithm predicts the length of the next CPU burst of each process?
6. (10%) Describe one advantage and one disadvantage of using large pages in the paging system.

Algorithms 資格考 2012 Spring

1. (30%) Solve $T(n) = 7T(\frac{n}{3}) + \Theta(n)$ using Θ .
2. (25%) Describe the steps to show a problem to be NP-complete.
3. (20%) Illustrate the operation of Counting sort on the array $A = (6, 0, 2, 0, 1, 3, 4, 6, 1, 3, 2)$.
4. (25%) Describe an algorithm that, given n integers in the range 0 to k , preprocesses its input and then answers any query about how many of the n integers fall into a range $[a..b]$ in $O(1)$ time. Your algorithm should use $\Theta(n+k)$ processing time.

VLSZ

101. 4. 13 **Name** _____ **St. No** _____

1. Explain the following terms in detail: (60%)

- (a) critical path (b) setup time
- (c) clock skew problem (d) Design for testability
- (e) hold time (f) soft IP

2. Suppose you have completed a circuit design with hardware description language. Now, you must decide to implement your circuit whether with (a) ASIC or (b) CPLD/FPGA. Which way is better? Explain the advantages and disadvantages of the two ways in detail. (20%)

3. Describe the difference between full custom and Cell-based design flow. (20%)

密碼學資格考 2012, 4,6 Close Book

- (1) 試論雲端計算上出現的重要資訊安全問題，並討論可能解決之道。(25%)
- (2) 何謂 Sequence of Games ? 它在資訊安全系統上有何用途？能否用於量子密碼上？請用BB84來說明。(25%)
- (3) 量子是什麼？有什麼重要的性質？在量子計算與資訊理論上有哪些重要的公設？(25%)
- (4) 請用因數分解為例子來說明量子計算的特色。(25%)

「無線通訊與行動網路」 資格考

[25%] 1. Please explain the function of Slotted ALOHA, CSMA non-persistent, CSMA/CD and CSMA/CA in pseudo-code format. Your description shall be detailed enough to convince me that you really understand the meaning of these three terms.

[10%] 2. We want to provide a wireless communication service with one phone sets on a long-distance mini bus carrying two passengers. Assume, on the average, each passenger makes a six-minute call in every hour. What is the probability of a passenger approaching the telephone and the phone line is busy? Given that the Erlang B Formula is $B(N, \alpha) = \frac{\alpha^N / N!}{\sum_{i=0}^N \alpha^i / i!}$.

[15%] 3. If a total of 50 MHz of bandwidth is allocated to a particular FDD cellular telephone system which uses two 10 KHZ simplex channels to provide full duplex voice and control channels, compute the number of channels available per cell if a system uses (a) three-cell reuse, (b) four-cell reuse, and (c) seven-cell reuse.

[10%] 4. Please derive the equation for the throughput of the pure Aloha protocol.

[20%] 5. Pretend your company won a license to build a U.S. cellular system. Your license is to cover 160 square km. Assume a base station costs \$50,000 and a core-network infrastructure costs \$1,500,000. An extra \$500,000 is needed to advertise and start the business. You have convinced the bank to loan you \$6 million, with the idea that in four years you will have earned \$9 million in gross billing revenues, and will have paid off the loan (It means totally you make \$15 million.).

- (a). How many base stations (i.e. cell sites) will you be able to install for \$6 million?
- (b). Assuming the earth is flat and subscribers are uniformly distributed on the ground, what assumption can you make about the coverage area of each of your cell sites? What is the major radius of each of your cells, assuming a hexagonal mosaic?
- (c). Assume that the average customer will pay \$50 per month over a four year period. Assume that on the first day you turn your system on, you have a certain number of customers which remains fixed throughout the year. On the first day of each new year, the number of customers using your system doubles and then remains fixed for the rest of that year. What is the minimum number of customers you must have on the first day of service in order to have earned \$ 9 million in gross billing revenues by the end of the 4th year of operation?
- (d). For your answer in (c), how many users per square km are needed on the first day of service in order to reach the \$ 9 million mark after the 4th year?

[20%] 6. Please give the system architecture of a 2nd generation (GSM) and GPRS telecommunication system. Draw the pictures to make your explanation more clearly. For the core-network part, please specify the hardware components that are commonly used in GSM and GPRS. And, what new components are introduced in the GPRS system?