

# Today's Lecture Logistics Course Outline Introduction



# Logistics



- Instructor: Mehdi Tahoori
  - Office: Room A.3.14, Building 07.21
  - Email: mehdi.tahoori@kit.edu
  - Tel: 721-608-47778, Fax: 721-608-43962Office hours: Wednesday 13:00-14:00
  - Secretary: Ms. Iris Schroeder-Piepka
- Lecture:
  - When: Wednesday 14:00-15:30Where: Room 236, Building 50.34
- Lecture notes
  - Available online: http://cdnc.itec.kit.edu
  - Under (Education → current semester → Reliable computing I)

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

3

# **Logistics (cont)**



- Requirements
  - Computer Architecture
- Background on (preferred but not required)
  - Logic Design
  - Algorithms and Programming
  - Operating system (OS)
  - Basic probabilities
- Related Lectures
  - Testing Digital Systems I and II

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



### **Reference Books**



- D.K. Pradhan, ed., Fault Tolerant Computer System Design, Prentice-Hall, 1996
- D.P. Siewiorek and R.S. Swarz, Reliable Computer Systems - Design and Evaluation, Digital Press, 1998, 3rd edition.
- I. Koren and C. M. Krishna, Fault-Tolerant Systems. Morgan Kaufmann, 2007
- P. K. Lala, Self-Checking and Fault-Tolerant Digital Design. Morgan Kaufmann, 2001
- B. W. Johnson, Design and Analysis of Fault Tolerant Digital Systems, Addison Wesley, 1989

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

5

#### **Course Outline**



- Introduction
  - Historical perspectives
  - Motivation for reliable system design
  - Failure sources
- Metrics and definitions
  - Defects, faults, errors, failures
  - Reliability metrics
  - Reliability evaluation
- Hardware reliability techniques
  - Error masking techniques
  - Error detection and recovery techniques
- Software reliability techniques

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



#### **Course Goals**



- Understanding the basic concepts in reliable system design, metrics, evaluation, and requirements
- Introduced to classical fault tolerant techniques
- Being familiar to current challenges in reliable computing
  - Maybe you become interested in doing research in reliable computing;-)
- Being able to apply some of these concepts to systems
  - Hardware and software

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

7

# **Course Outline (detailed)**



- Lecture 1: Introduction to reliable computing
- Lecture 2: Reliability metrics
- Lecture 3: Faults, errors, failures
- Lecture 4: Hardware redundancy
- Lecture 5: Reliability evaluation
- Lecture 6: Information redundancy-1
- Lecture 7: Information redundancy-2
- Lecture 8: Redundant disk arrays
- Lecture 9: Concurrent Error Detection
- Lecture 10: Re-execusion
- Lecture 11: Checkpointing and Recovery
- Lecture 12: Software fault tolerance

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# **Today's lecture: Outline**



- Need for reliability: real disaster stories!
- Reliability: past, present, future
- Challenges
- Examples of reliable system design

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

# Correct operation in computing These are the system components. All are needed for proper operation Good software Correct human operation (c) 2017, Mehdi Tahoori Reliable Computing 1: Lecture 1



#### **Motivation**



- Fault tolerance has always been around
  - NASA's deep space probes
  - Medical computing devices (e.g., pacemakers)
  - But this had been a niche market until fairly recently
- But now fault tolerance is becoming more important
  - More reliance on computers
- Extreme fault tolerance
  - Avionics, car controllers (e.g., anti-lock brakes), power plants and power delivery network, medical systems, etc.
- High fault tolerance
  - Commercial servers (databases, web servers), file servers, high performance computing (HPC), etc.
- Some fault tolerance
  - Desktops, laptops (really!), smartphones, game consoles, etc.

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

11

# Why We Need High Reliability?



- High availability systems:
  - Telephone
  - Transaction processing: banks/airlines
- Long life missions:
  - Unscheduled maintenance too costly
  - Long outages, manual reconfiguration OK
  - Critical applications
- Critical applications:
  - Real-time industrial control
  - Flight control
- Ordinary but widespread applications:
  - CDs: encoding
  - Internet: packet retransmission

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# **Stuff Happens**



- We wouldn't need fault tolerance otherwise!
- Physical problems
  - Melted wire
  - Toasted chip
- Design flaws
  - Incorrect logic (e.g., Pentium's FDIV, AMD's quad-core TLB bug)
  - Buggy software (e.g., Vista)
- Operator error
  - Incorrect software installation
  - Accidental use of rm R \*
- Malicious attacks
  - Security is beyond the scope of this course

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

12

# eBay Crash



- eBay: giant internet auction house
  - A top 10 internet business
  - Market value of \$22 billion
  - 3.8 million users as of March 1999
  - Bidding allowed 24x7
- June 6, 1999
  - eBay system is unavailable for 22 hours with problems ongoing for several days
  - Stock drops by 6.5%, \$3-5 billion lost revenues
  - Problems blamed on Sun server software
- Shorter downtimes common

(c) 2017. Mehdi Tahoori

Reliable Computing I: Lecture 1



#### **Ariane 5 Rocket Crash**



- Ariane 5 and its payload destroyed about 40 seconds after liftoff (1996)
- Error due to software bug:
  - Conversion of floating point to 16-bit int
  - Out of range error generated but not handled
- Testing of full system under actual conditions not done due to budget limits
- Estimated cost: 120 million DM

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

15

#### The Therac-25 Failure



- Therac-25 is a linear accelerator used for radiation therapy
- More dependent on software for safety than predecessors (Therac-20, Therac-6)
- Machine reliably treated thousands of patients, but occasionally there were serious accidents, involving major injuries and 1 death (1985-1987).
- Software problems:
  - No locks on shared variables (race conditions).
  - Timing sensitivity in user interface.
  - Wrap-around on counters.

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



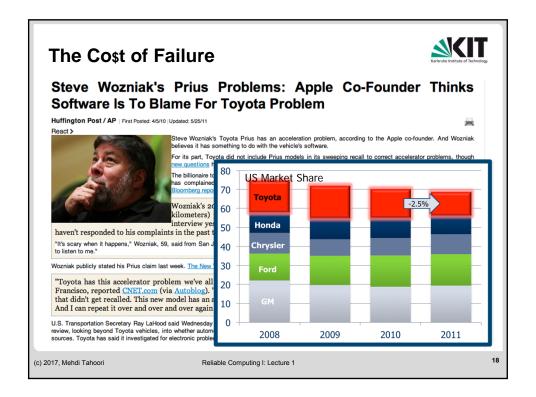
#### **Tele Denmark**



- Tele Denmark Internet, ISP
- August 31, 1999
  - Internet service down for 3 hours
  - Truck drove into the power supply cabinet at Tele Denmark
  - Where were the UPSs?
    - Old ones had been disconnected for upgrade
    - New ones were on the truck!

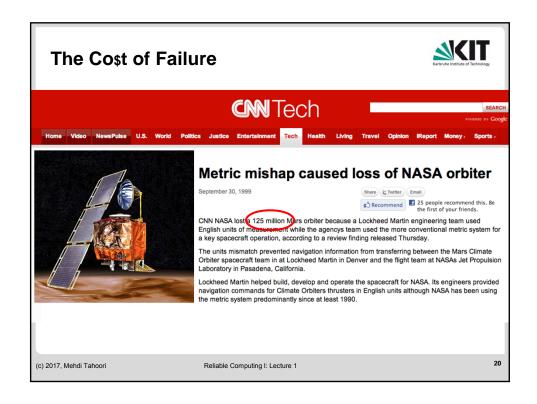
(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1









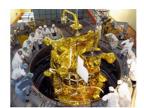


#### The Cost of Failure



Weekly Inside Technology News

February 23, 2012



# Did Bad Memory Chips Down Russia's Mars Probe?

The failure of Russia's ambitious Phobos-Grunt space mission has been shrouded in confusion and mystery—from the first inklings of trouble after its November launch to inconsistent reports of where it fell to Earth in January, and beyond. The release of the official accident investigation results on 3 February served only to further rumors of fundamental hardware and software design flaws, and of blatant violations of safety standards. The report blames the loss of the probe primarily on memory chips that became fatally damaged by cosmic rays. But observers place the blame squarely on the engineers who put Phobos-Grunt together, noting that if these military-grade chips (which were not radiation hardened) had been proposed for a critical component in a space-probe design for the U.S. space program, they probably would not have been approved for use. Worse still for Russia's space agency are reports suggesting that the chips were counterfeits that had been intentionally misrepresented as offering higher performance than they were actually capable of.

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

21

#### The Cost of Failure



#### The Pentium Problem

Late last year [1994] there was a major flap in the media about Intel's Pentium (TM) chip.

Intel publicly announced that "an error is only likely to occur [about] once in nine billion random floating point divides", and that "an average spreadsheet user could encounter this subtle flaw once in every 27,000 years of use." Critics noted that while hitting a pair of "had inputs" was unlikely, the Pentium's output for those inputs was wrong every time. Others suggested that some "bad inputs" might occur with disproportionate frequency in common calculations. Many noted that without completely repeating massive calculations on other computers, they could never tell if they had indeed encountered any of the bad inputs. Within a month IBM halted shipment on Pentium-based computers (which comprised only a small percentage of IBM's computer production) and announced that "Common spreadsheet programs, recalculating for 15 minutes a day, could produce Pentium-related errors as often as once every 24 days."

Intel's policy, when it first publicly admitted the problem around November 28 of 1994, was to replace Pentium chips only for those who could explain their need of high accuracy in complex calculations. (Being a math professor seemed to help.) Great public outcry ensued, with Intel the butt of many jokes. By late December Intel capitulated and announced a free replacement Pentium for any owner who asked for one.

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



#### The Cost of Failure



- Samsung Galaxy Note 7 exploding
  - Samsung's expected losses from the Galaxy Note 7 catastrophe have soared above \$5 billion.
  - More than \$19bn was wiped off the company's market value





(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

23

#### **SKIT Examples of Computer-related Failures** FAILURES onfidentiality Availability Reliability Design False alerts at the North American Air Defense June 1980 (NORAD) [Ford 85] √ First launch of the Space April 1981 Shuttle postponed [Gaman 81] √ √ Excessive radiotherapy doses June 1985 - January 1987 (Therac-25) [Leveson & Turner 93] 1 1 November 1988 Internet worm [Spatford 89] $\sqrt{}$ 9 hours outage of the long-distance 15 January 1990 phone in the USA [Neumann 95] √ 1 Scud missed by a Patriot February 1991 (Dhahran, Gulf War) [Neumann 95] √ Crash of the communication system of November 1992 the London ambulance service [HA 93] $\sqrt{}$ 26 and 27 June 1993 Authorization denial of credit card operations in France 1 $\checkmark$ $\checkmark$ $\sqrt{}$ 4 June 1996 The maiden flight of the Arine 5 launcher ende √ √ in a failure (France) July 20, 2008 8-hour outage of Amazon services Reliable Computing I: Lecture 1 (c) 2017, Mehdi Tahoori



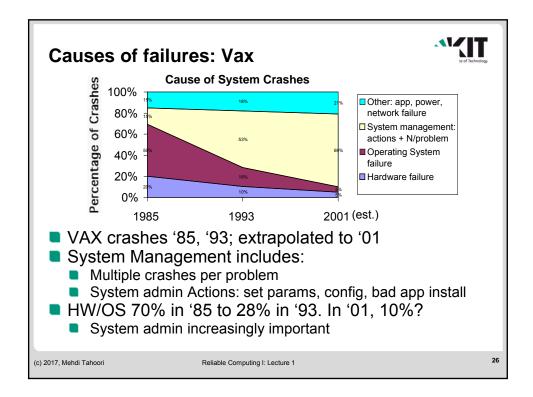
#### Causes of failures: Tandem



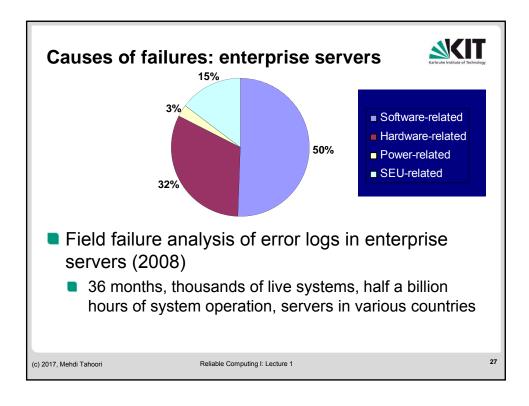
- Dominant manufacturer of fault-tolerant computer systems
  - For ATM networks, banks, stock exchanges, telephone switching centers, and other similar commercial transaction processing applications
  - Now part of HP
- In Gray's '85 survey of Tandem customers
  - 30% were "infantile" failures
  - The rest were broken into (roughly):
    - Administration 42%
    - Software 25%
    - Hardware 18%
    - Environment (power, etc.) 14%
    - Unknown 3%

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1







#### Costs!!!



Example of Average Cost per Hour of Downtime

Industry	Application	Average Cost per Hour of Downtime
Financial	Credit Card Sales	\$2,600,000
Media	Pay-per-view	\$1,150,000
On-line commerce	Ebay	\$225,000
Retail	Catalog Sales	\$90,000
Transportation	Airline Reservation	\$89,500

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# Reliability: increasing concern



- Historical
  - High reliability in computers was needed in critical applications: space missions, telephone switching, process control etc.
- Contemporary
  - Extraordinary dependence on computers: on-line banking, commerce, cars, planes, communications etc.
  - Hardware is increasingly more fault-prone
  - Software is increasingly more complex
  - Things simply will not work without special reliability measures

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

20

### Why Study Reliable Computing!!!



- Traditional needs
  - Long-life applications (e.g., unmanned and manned space missions)
  - Life-critical, short-term applications (e.g., aircraft engine control, fly-by-wire)
  - Defense applications (e.g., aircraft, guidance & control)
  - Nuclear industry

# Newer critical-computation applications

- Health industry
- Automotive industry
- Industrial control systems, production lines
- Banking, reservations, switching, commerce

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# Why Study Reliable Computing!!! (cont.)



- Networks
  - Wired and wireless networked applications
  - Data mining
  - Information processing on the Internet
  - Distributed, networked systems (reliability and security are the major concerns)
  - Intranet stores, catalog industry (commercial computing)
  - Cloud computing
- Scientific computing, education
  - Arrival of supercomputers puts high requirements on reliability
  - Petascale and exascale computing

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

31

#### Need for fault tolerance: Universal & Basic



- Natural objects:
  - Fat deposits in body: survival in famines
  - Clotting of blood: self repair
  - Duplication of eyes: graceful degradation upon failure
- Man-made objects
  - Redundancy in ordinary text
  - Asking for password twice during initial set-up
  - Duplicate tires in trucks
  - Coin op machines: check for bad coins

(c) 2017. Mehdi Tahoori

Reliable Computing I: Lecture 1



#### Reliability issues are not new...

■ In the July 1834 issue of the Edinburgh Review, Dr. Dionysius Lardner published the article

"Babbages's calculating engine", in which he wrote:





"The most certain and effectual check upon errors which arise in the process of computation, is to cause the same computations to be made by separate and independent computers; and this check is rendered still more decisive if they make their computations by different methods."

Babbage's Calculating Engine (1833)

(c) 2017, Mehdi Tahoori

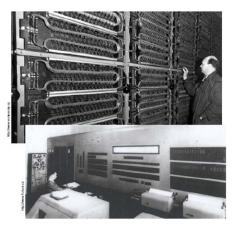
Reliable Computing I: Lecture 1

33

#### **First Fault-Tolerant Computer**



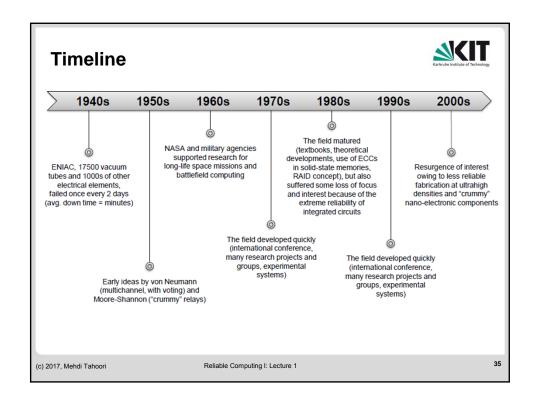
The SAPO (Samočinný počítač) fault-tolerant computer system

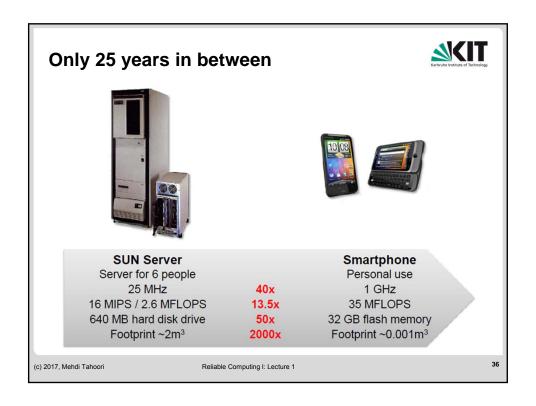


- Designed and built between 1950 and 1956 by Antonín Svoboda in Prague
- Three parallel arithmetic units, which decided on the correct result by voting (TMR)
- Electromechanical design:
  - 7000 relays,
  - 400 vacuum tubes
  - magnetic drum memory
- The system burnt down in the year 1960 after a relay failure

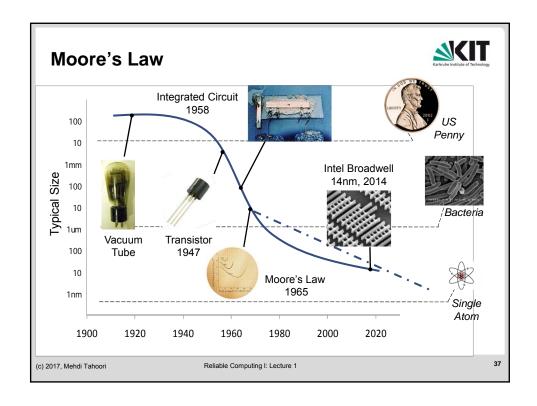
(c) 2017, Mehdi Tahoori

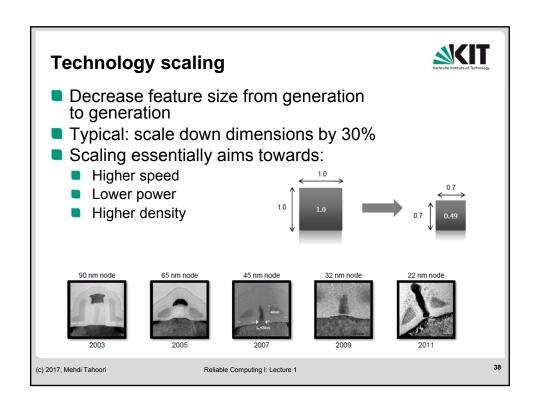




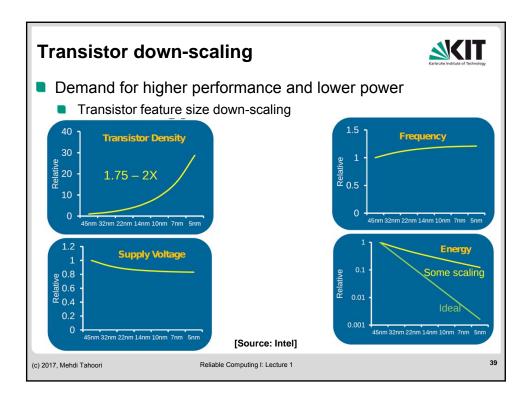


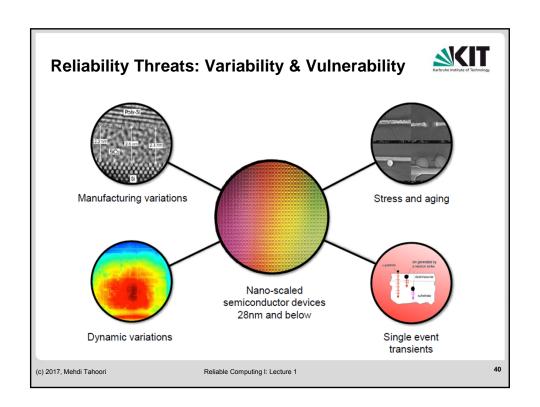




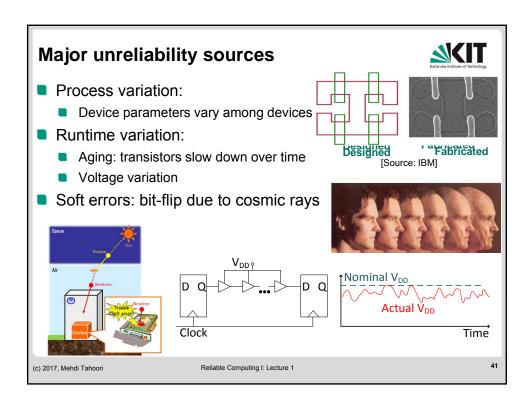


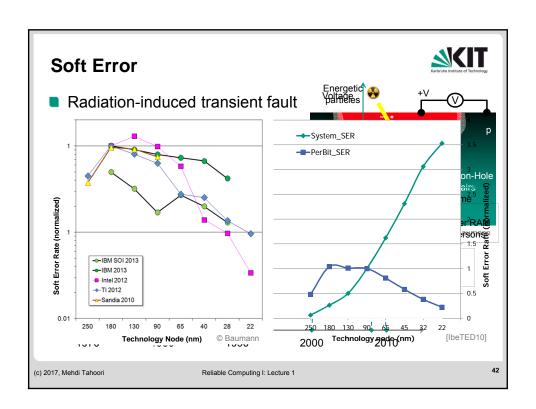




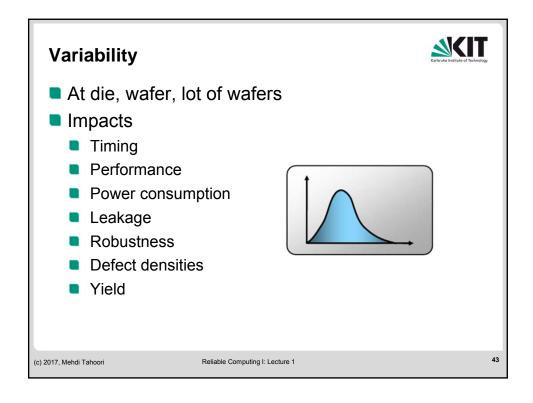


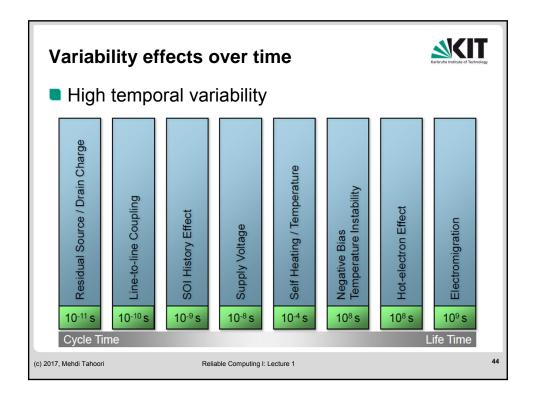






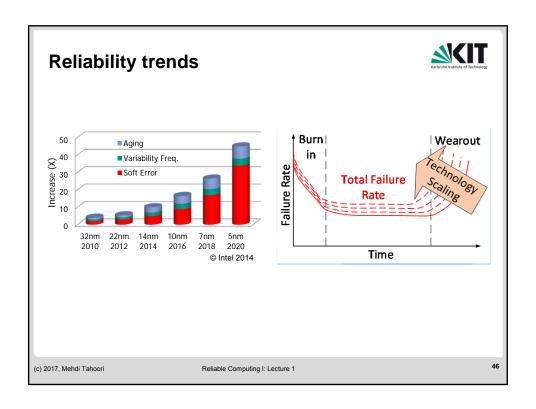




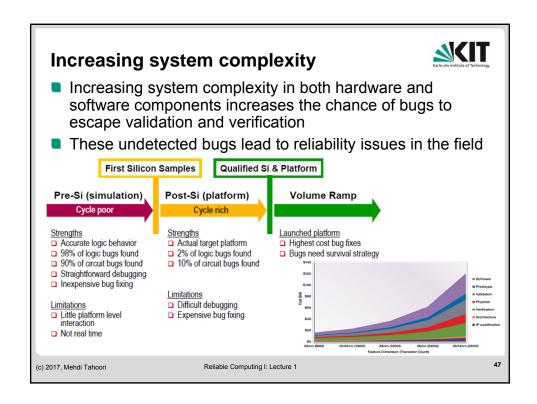


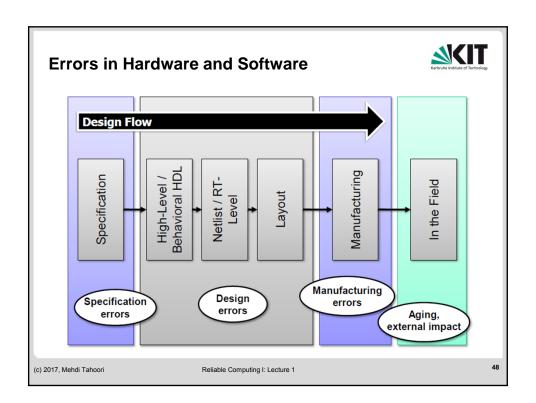


# Increasing Device Count per Chip Increased number of transistors that each chip statistically samples from the device parameter distribution. To achieve comparable chip-level yields via margining, we are forced to accept a larger spread of device characteristics.











# Why Fault Tolerance Isn't Easy



- Fault tolerance can be solved to any arbitrary degree if you're willing to throw resources at the problem
- Resources to sacrifice:
  - System performance
  - Cost
  - Power

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

40

# **Trying Not To Lose Performance**



- There are many FT approaches that sacrifice performance to tolerate faults
- Example 1
  - Periodically stop the system and checkpoint its state to disk
  - If fault occurs, recover state from checkpoint and resume
- Example 2
  - Log all changes made to system state in case recovery is needed
  - During recovery, undo the changes from the log
- Example 3
  - Run two identical systems in parallel
  - Compare their results before using them
- Example 4
  - Run software with lots of assertions and error checking

(c) 2017. Mehdi Tahoori

Reliable Computing I: Lecture 1



#### **Performance Issues**



- Most important not to degrade performance during fault-free operation
  - This is the common-case → make it fast!
  - Amdahl's Law
- Somewhat less important not to degrade performance when a fault occurs
  - This still might not be acceptable in certain situations (e.g., real time systems)

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

51

# **Trying Not To Increase Cost**



- There are many FT approaches that sacrifice cost to tolerate faults
- Example 1
  - Replicate the hardware 3 times and vote to determine correct output
- Example 2
  - Mirror the disks (RAID-1) to tolerate disk failures
- Example 3
  - Use multiple independent versions of software to tolerate bugs
    - Called N-version programming

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# **Trying Not To Increase Power**



- There are many FT approaches that sacrifice power to tolerate faults
- Examples 1, 2 & 3 (same as previous slide)
  - Replicate the hardware 3 times and vote to determine correct output
  - Mirror the disks (RAID-1) to tolerate disk failures
  - Use multiple independent versions of software to tolerate bugs
- Example 4
  - Add continuously running checking hardware to system
- Example 5
  - Add extra code to check for software faults

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

53

#### **Levels of Fault Tolerance**



- Fault tolerance can be at many levels in a system:
- Application software
  - Adding assertions to code
- Operating system
  - Protecting OS from hanging
- Entire hardware system
- Hardware sub-system
- Circuits and transistors

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# **Example 1: Telephone Switching System**



- Extreme availability
  - Goal: <3 minutes of downtime per year</p>
  - Goal: <0.01% of calls processed incorrectly</p>
- Uses physical redundancy
- Hardware cost: about 2.5 times cost of equivalent non-redundant system
- Also uses:
  - Error detecting/correcting codes (e.g., parity, CRC)
  - Watchdog timers
  - Many forms of diagnostics
  - Dynamic verification ("sanity program")

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

55

# **Example 2: IBM Mainframes**



- Lots of fault tolerance → high availability
- Note: IBM has produced mainframes since the 1960s, and they've changed their design and enhanced their fault tolerance several times since then
- Redundancy at many levels
  - Redundant units within processor (e.g., register file)
  - Redundant processors
- Diagnostic hardware for isolating faults
- Reliable operating system

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1



# **Example 3: My (previous) Laptop**



- Minimal fault tolerance
- Designed to be cheap and fast ... and obsolete in a few years
- May have parity or ECC on:
  - Some bus lines
  - DRAM
  - Hard disk
- Expected lifetime (as expected by me): 2 years
- Expected Mean Time To Reboot (also by me): 1 week
- Expected Mean Time To Re-install: 6 months
- Expected probability of it successfully coming out of "hibernation" mode: 0

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

57

#### **Example 4: Database Software (Oracle, DB2)**

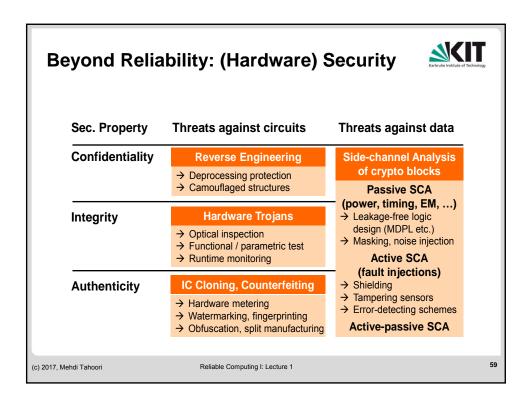


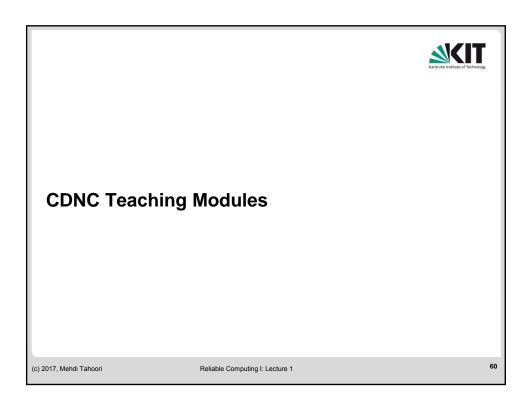
- Lots of fault tolerance
- Can't afford to corrupt vital data
- Enforces "ACID" properties for transactions & data
  - Atomicity: transactions are atomic
  - Consistency: only valid data written to database
  - Isolation: transactions don't interfere with each other
  - Durability: data written to database won't be lost
- Implemented with logging and checkpointing
- Writes important data to disks
  - Can even tolerate a fault that occurs while writing to disks

(c) 2017, Mehdi Tahoori

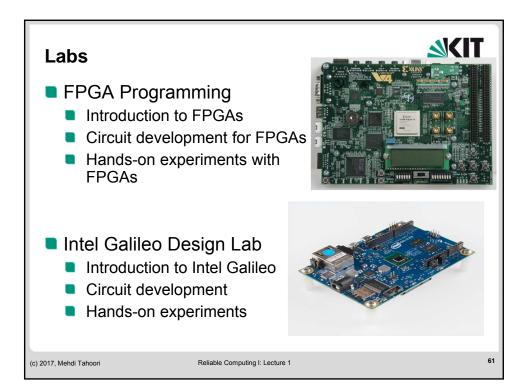
Reliable Computing I: Lecture 1

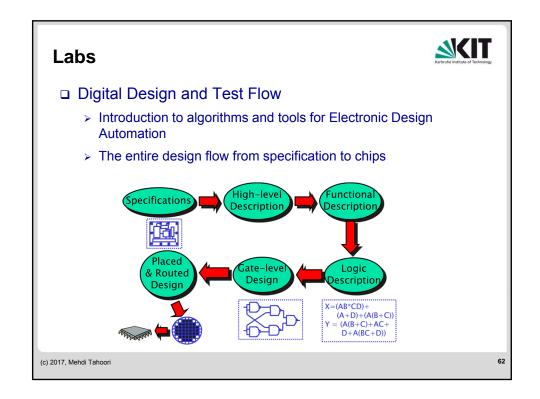














#### **Seminars**



- Non-Volatile Memory Technologies
  - Introduction to non-volatile memories
  - Can these be used to replace SRAM/DRAM?
- Near Threshold Computing
  - A standard transistor operates with a voltage >> Vth
  - In near threshold computing the voltage is lowered to ~Vth
  - Advantage: Huge power savings
  - Challenge: Performance

(c) 2017, Mehdi Tahoori

Reliable Computing I: Lecture 1

63

#### HiWi and Studienarbeit



- Hiwi positions
  - > To help with various research projects in dependable computing
  - > Required knowledge
    - Programming
    - Digital design and computer architecture (DT+RO)
- Studienarbeit
  - > Various projects related to
    - · Computer archiecture and logic design
    - · Dependable computing

Contact us if interested & for more details: cdnc@ira.uka.de

(c) 2017, Mehdi Tahoori