

Bits, Bytes, and Data Abstraction

Introduction

Modern computers only use zeroes and ones for storing and processing data. How do computers produce pictures, video, and sound using only ones and zeroes?!

In this lesson you will develop programs on an Android™ device. Some of the data (namely, colors) will require the use of numbers that range from 0 to 255. Why is 255 the largest number used?



MATERIALS

- Lego bricks for each student

Procedure

Part I. Represent and store numbers and letters in binary

Refer to your downloadable resources for this material. Interactive content may not be available in the PDF edition of this course.

1. Form pairs as directed by your teacher.
2. Instead of the decimal number system, computers like the Android devices you'll use in the next activities store and process most numbers using the binary number system, which is base two. The decimal number system you are familiar with utilizes ten distinct **digits**, 0 through 9, and it has place values that are powers of 10. We call this the base ten number system.

Example digits:		9	0	4	7
Place values:	10^n	1000	100	10	1
Powers of ten:	10^n	10^3	10^2	10^1	10^0

Explain why the sequence of digits 9047 represents “nine thousand forty seven.”

3. A binary number is represented using only two digits: 0 and 1. A single zero or one is called a **bit**, a humorous abbreviation for “binary digit.” The place values in the binary number system are powers of 2. What are those place values?

Example digits:		1	0	1	1	1	0	1	1
Place values:		—	—	—	—	—	4	2	1
Powers of 2:	2^n	—	—	—	—	—	2^2	2^1	2^0

4. To have a hands-on experience with the **binary** number system, you will create a “device” like a hard drive, but it will only store a combination of eight zeroes or ones. Eight bits are called a **byte**. The byte is a binary representation of a number between 0 and 255 from the decimal number system. The place values of each bit within a byte are shown below. Each brick will store one bit—as a single zero or one. For example, a byte with a value of 00010001 (in binary) represents the value 17 (in decimal) because the ones are in the 1s place and the 16s place. Since binary has only two digits, you can think of the place values with a 1 as being *on* and the place values with a 0 as being *off*.

Place Value		Binary Bit		Decimal Value
128	*	0	=	0
64	*	0	=	0
32	*	0	=	0
16	*	1	=	16
8	*	0	=	0
4	*	0	=	0
2	*	0	=	0
1	*	1	=	1
		Total	=	17

- What is the decimal value of 00001111?
- What is the binary representation for decimal value of 22?
- Take a few moments to practice converting between binary and decimal representations. Visit <http://www.mathwarehouse.com/non-decimal-bases/> and play the "Cisco Binary Numbers Game"

5. The meaning of the value of a byte depends on what it represents. When you develop Android applications, you will represent colors, numbers, and characters. The number could be the red, green, and blue intensities of a color on the screen, or it could be the position of a speaker cone that generates sound. To represent letters, a common code is used: the **ASCII code**. To accommodate letters from other alphabets, usually we use the **UTF-8** code. UTF-8 uses all the values from ASCII and then extends it with additional bytes to include other characters. The ASCII and UTF-8 codes are easy to find online. Here is part of that code:

Letter:	ASCII Code:	Binary:
A	65	01000001
B	66	01000010
D	?	?
Z	90	?

a	?	?
b	98	?
z	122	1111010

Based on this table, what ASCII or UTF-8 value encodes for the character D?

How would you write that in binary?

6. Your teacher might provide you with a notecard and some rectangular building blocks. Gather these materials as directed by your instructor. Position each block on the notecard in the face down positions indicated in the image below.



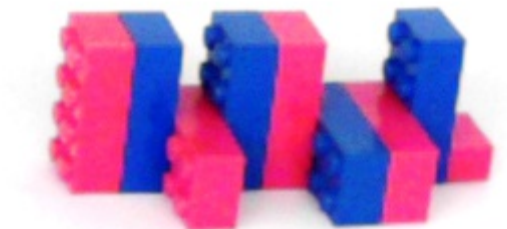
7. Record and read a byte following these steps:
 - To record a bit, flip each building block to the appropriate position:
 - Face down block = 0 (blocks in starting position as shown in the previous step)
 - Block standing on end = 1 (blocks on end as in the image below)



Either you or your partner will record values to send to the other person. If you are the first to record, record a 1 or a 0 for each building block until you have recorded a known value for all 8 bits. To decide what value to record, send your partner a secret message of one ASCII letter recorded in binary, perhaps the first letter of your answer to a question your partner asks, like “What is your favorite animal?” or “What is your favorite class?” Secretly record your plan for 0s and 1s here. For example, B = 66 = 01000010

0s and 1s:	—	—	—	—	—	—	—	—
Place values:	128	—	—	—	—	—	2	1

- To read the value of the bits stored by your partner, look at the notecard. From left to right, record a 0 if a block is face down and a 1 if a block is standing up. For example the image shown below records the binary representation 11011001.

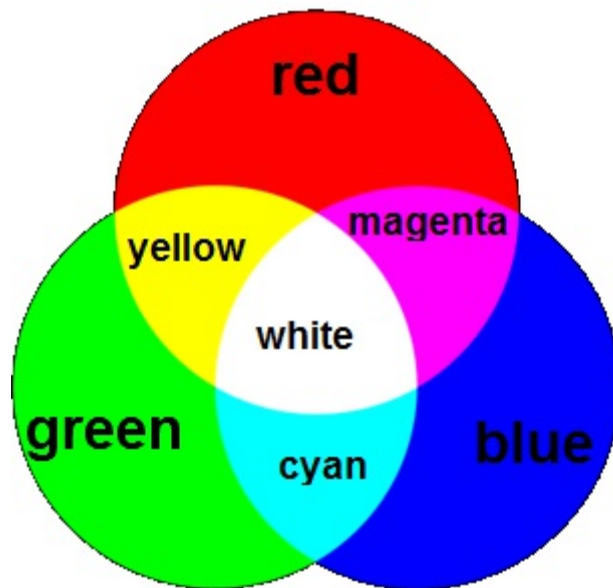


- After you record a byte and your partner reads it, switch roles.

Part II. Represent images

8. Bits can represent pictures as well as text. A graphic image is represented with **pixels**. For example, the sprites that you animated in Scratch™ programming language were all bitmapped images with distinct values associated with each pixel. You may choose to use such images again in App Inventor.

In one representational format (RGB, 3 bytes/pixel), the color of each pixel is stored by representing the intensity of red with the first byte, green with the second byte, and blue with the third byte. All color perception by humans can be represented by mixing these three colors of light. The diagram below shows the mixing needed to obtain secondary additive colors; these are different than the secondary subtractive colors you are familiar with. Magenta, for example, is a mixture of red and blue light and is represented by 255, 0, 255.



- What color is 255,0,0?
 - What color is 0,255,255?
 - Since yellow is red + green, it is represented by 255, 255, 0. The color can be made more greenish by reducing the red, such as 200, 255, 0. The color can be made dimmer by reducing all numbers, such as 100, 127, 0. What values would represent a very dark blue with a slightly reddish tint?
9. Image files contain the RGB values for all pixels and meta-information like the author and date, data format, or camera settings. Most image formats also compress the data as described below. Ignoring meta-information and compression, the image file size is easy to calculate. A 200 x 300 pixel image with 3 bytes/pixel is $200 \times 300 \times 3 \text{ bytes} = 180000 \text{ bytes} = 180 \text{ kb}$.

What is the file size of an image that is 600 x 800 pixels with 3 bytes/pixel?

10. Video is presented as a series of frames, each frame representing an image. One common standard (like MP4) to represent a movie uses 640 x 360 pixels for each frame and uses 24 frames/second. With 3 bytes/pixel, what is the file size of a 1 minute video?

Part III. Describe data compression

11. Images and other files are often compressed to make the files smaller. This is the case with the APK source code files that you will send between your computer and your Android device. Compression is useful whenever you transfer data between devices or over a network because it reduces the amount of time that it takes to send and receive the file. One compression scheme takes advantage of repeated numbers:

1 1 1 2 2 2 2 2 7 6 is compressed to 3 1 6 2 1 7 1 6. The compressed format indicates that there are three 1s, six 2s, and so on.

- What is the compressed representation of 5 5 5 5 5 5 5 5 3 3 3?
- The **compression ratio** tells how much a file is shrunk by **compression**. There are two ways to report a compression ratio that are equally acceptable but not equivalent. One reports the percentage of the file size that can be eliminated; the other reports the ratio of uncompressed to compressed sizes. Either way, a higher compression ratio means more compression.

$$\text{Percent Compression} = 1 - \frac{\text{Compressed File Size}}{\text{Uncompressed File Size}}$$

OR

$$\text{Compression Ratio} = \frac{\text{Uncompressed File Size}}{\text{Compressed File Size}}$$

If each number in the example is stored in one byte, then there are 11 bytes being compressed to 8 bytes.

The percent compression was $1 - 8/11 = 100\% - 72\% = 28\%$.

The compression ratio was $11/8 = 1.375:1$. These are not equivalent, but both methods are used.

What percent compression was achieved in the example in part a? What compression ratio?

12. The compression method in question 11 was lossless. The resulting data was more compactly represented, but all information is still there. Most compression algorithms (like JPG) are **lossy**, meaning that some information has been left out or lost and the original data cannot be

re-created from the compressed data. One lossy compression method uses a limited color **palette** so that the color of each pixel can be specified with fewer than 3 bytes. This method will create large blocks of uniform color, losing the slight variations in the original image.

- The greater the lossiness, the smaller the file (good) but the lower the **fidelity** (potentially bad). For images and video, low fidelity means low quality. Companies transmitting via Internet have to balance quality of the video against quality of the transmission speed. Think of an Internet video you have seen and decide whether you would prefer higher fidelity or higher transmission speed and explain why.
- In video conferencing software, face detection algorithms are used to identify portions of each frame where lower compression ratios should be used: the portions containing faces. Explain the benefits of using this algorithm in terms of both fidelity and speed.

Part IV. Describe connection to society. Distinguish analog vs. digital

There are many examples of software failures throughout history. Knowing how and why some of these errors occur will help you to develop more functional applications.

13. Bit representations of numbers are limited to a maximum value for a certain number of bits.



In 1996 the European Space Agency had to detonate an Ariane 5 rocket and its spacecraft payload because of software design. In its guidance software, a 64-bit number representing acceleration was truncated to a 16-bit number. The software was inherited from the Ariane 4 rocket, which didn't accelerate as much. For more information, see http://en.wikipedia.org/wiki/Ariane_5#Notable_launches.

What is the maximum value for a 64-bit number?

What is the maximum value for a 16-bit number?

Hint: Your answer depends on how the bits represent the number.

14. **(Optional)** Like any place value system, the binary number system cannot exactly represent some rational numbers. Rational numbers can be written as a fraction of two integers. They either terminate or repeat in any number system with place values, and those that repeat cannot be stored in a place value system without rounding.



Reduced fractions are repeating decimals if their denominator (when the fraction is reduced) has prime factors other than 2 and 5. This is because only 2 and 5 are the prime factors of the base 10.

In 1991, the United States used Patriot missiles to defend against the Iraq military's SCUD missiles in the U.S.-Iraq war. Unfortunately, the Patriot missile launchers did not work correctly. Among the causes was a software design flaw related to a timer for sending radar detection bursts. The timer was incremented in 0.1 second increments, with the accumulated value stored in – of course – binary. Because $1/10$ has a denominator with prime factors other than 2, this is a repeating binary number. The software truncated the representation, and the small error accumulated over hours of operation. After the design flaw was discovered, U.S. Army soldiers kept the Patriot missile launchers powered off when not being deployed so that the error would not accumulate. One launcher that was left powered on accumulated an error of 0.3 second before being used to defend against an incoming SCUD missile. The Patriot missiles it launched were aimed 0.3 seconds behind the actual position of the incoming SCUD missile. The SCUD killed 28 American soldiers when it struck their barracks.

- Give examples of rational numbers between 0 and 1 that can be represented as decimal without repeating and other examples that repeat.
- Give examples of rational numbers between 0 and 1 that can be represented in binary without repeating and other examples that repeat.

15. The acceleration and time mentioned in the previous two examples are continuous quantities: instead of jumping from one value to another, continuous quantities vary smoothly, having all intermediate values along the way. Continuous values can only be represented by

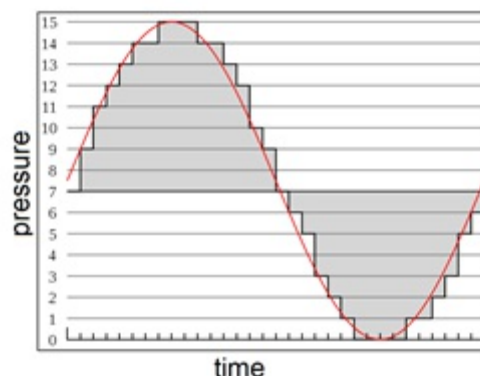
analog data. Analog data stores all possible continuous values within a range, like a needle on a meter.



The opposite of continuous is discrete. Discrete quantities jump from one value to another without ever equaling the values in between. Discrete value is represented by digital data. Digital data can only have specific values within a range. The examples illustrate why using digital data to describe analog quantities can be problematic.

Physical quantities like acceleration and time are often continuous quantities. The sounds that we experience, including music, are actually air pressure quickly increasing and decreasing. Air pressure is continuous, and analog recordings like vinyl records or cassette tape represent sound using continuous data like vinyl thickness or magnetic field strength.

With the widespread release of audio CDs in 1983, the music industry began a fast transition from analog to digital. The figure at right shows analog-to-digital conversion from analog (red) to 4-bit digital (gray). How you can tell that the digital representation requires 4 bits per pressure level?



16. When we convert continuous quantities to digital data, we lose information. However, digital data have huge advantages. Digital measurements can often collect data with higher

precision, and digital data can be copied over and over without errors accumulating. Digital data can also be duplicated and distributed at very low cost. Does this encourage or inhibit society's production and distribution of music and more generally creative ideas and expression? It is a controversial issue. Try to make the case for each side of this issue.

Part V. Describe data abstraction

17. When a computer program stores data, it usually also stores the **type** of data so that the computer knows what the zeroes and ones represent. A programmer often wants to name a color for a pixel, for example, without worrying about how the color is represented in zeroes and ones, and certainly without worrying about the magnetic, optical, or electronic phenomena representing the zeroes and ones. As a computer user, you have probably moved image files from one place to another without knowing anything about the individual pixels in the image. The image is stored using details stacked one level on top of another: details about magnetic fields, 0s and 1s, pixels, and compression. Separating the different levels of detail representing data is called **data abstraction**.
 - Describe the data abstraction that might be associated with the data in an audio file.
 - Explain why computers and computation wouldn't be practical without data abstraction.

Refer to your downloadable resources for this material. Interactive content may not be available in the PDF edition of this course.

Conclusion

1. Describe information that can be represented by zeroes and ones.
2. What is the relationship among data compression, data transmission speed, and fidelity vs. lossiness?
3. Consider the music produced by a band. Describe how both the terms “digital” and “analog” can be applied to the music produced and recorded by a band.