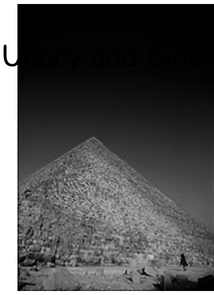
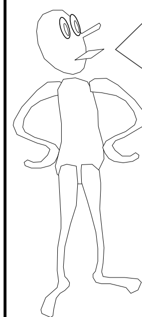
Great Theoretical Ideas In Computer Science		
Anupam Gupta		CS 15-251 Fall 2005
Lecture 4	Sept 8, 2005	Carnegie Mellon University

Unary and Binary





Your Ancient Heritage

Let's take a historical view on abstract representations.

Mathematical Prehistory: 30,000 BC

Paleolithic peoples in Europe record unary numbers on bones.

1 represented by 1 mark
2 represented by 2 marks
3 represented by 3 marks
4 represented by 4 marks
…

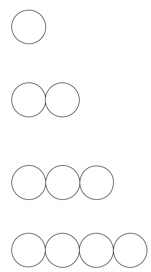
Prehistoric Unary

1

2

3

4



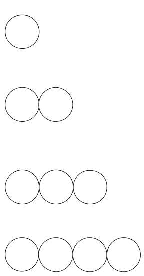
PowerPoint Unary

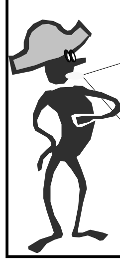
1

2

3

4



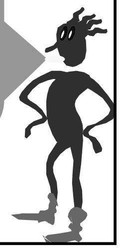


Hang on a minute!

Isn't unary a bit *literal* as a representation? Does it deserve to be viewed as an "abstract" representation?

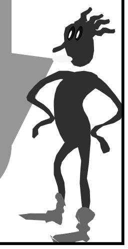
In fact, it is important to respect the status of each representation, no matter how primitive.

Unary is a perfect object lesson.



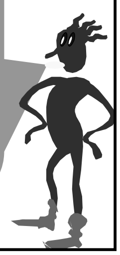
Consider the problem of finding a formula for the sum of the first n numbers.

We already used induction to verify that the answer is ½n(n+1)



Consider the problem of finding a formula for the sum of the first n numbers.

First, we will give the standard high school algebra proof…



1 + 2 + 3 + . . . + n-1 + n = S

n + n-1 + n-2 + . . . + 2 + 1 = S

(n+1) + (n+1) + (n+1) + . . . + (n+1) + (n+1) = 2S

n (n+1) = 2S

$$S = \frac{n(n+1)}{2}$$

1 + 2 + 3 + . . . + n-1 + n = S

n + n-1 + n-2 + . . . + 2 + 1 = S

(n+1) + (n+1) + (n+1) + . . . + (n+1) + (n+1) = 2S

n (n+1) = 2S

Algebraic argument

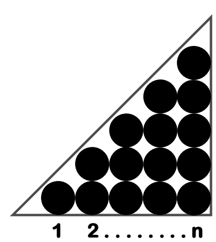
Let's restate this argument using a UNARY representation

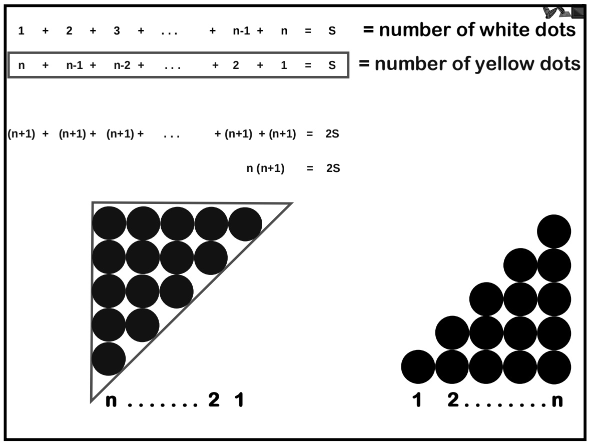
1 + 2 + 3 + . . . + n-1 + n = S **= number of white dots.**

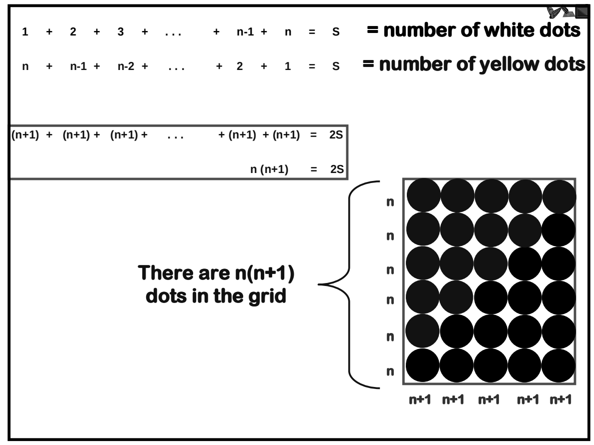
n + n-1 + n-2 + . . . + 2 + 1 = S

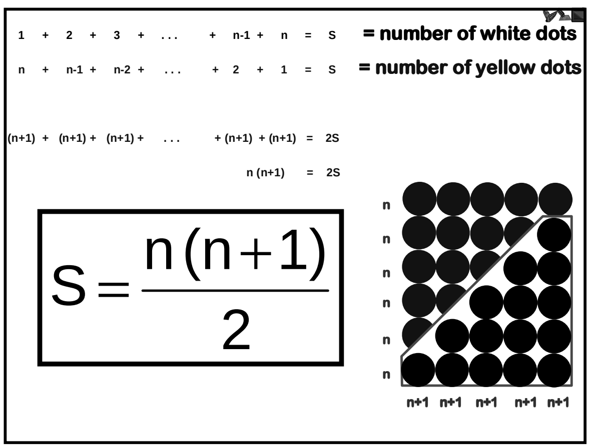
(n+1) + (n+1) + (n+1) + . . . + (n+1) + (n+1) = 2S

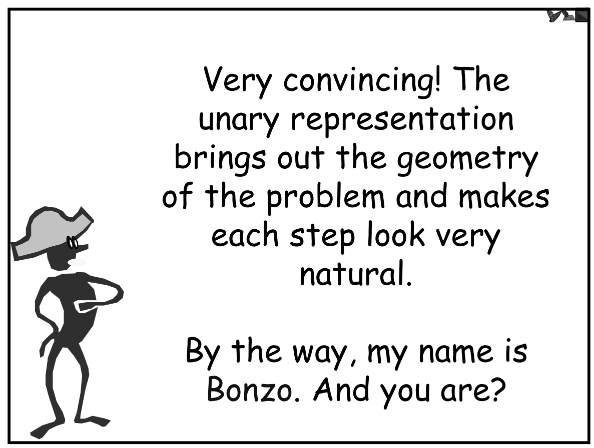
n (n+1) = 2S



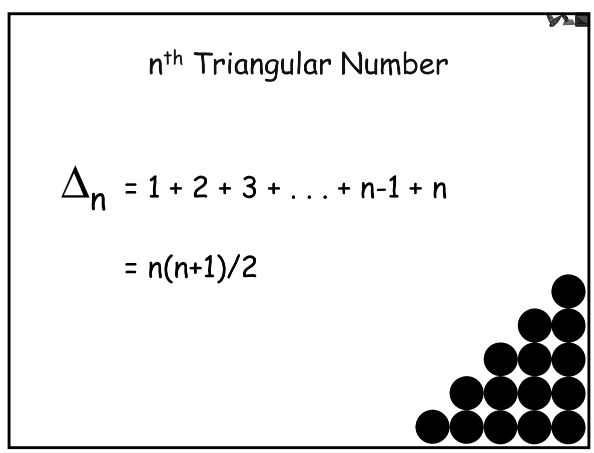

1 + 2 + 3 + ... + n-1 + n = S
= number of white dots
n + n-1 + n-2 + ... + 2 + 1 = S
= number of yellow dots
(n+1) + (n+1) + (n+1) + ... + (n+1) + (n+1) = 2S
n (n+1) = 2S
n 2 1
1 2 . . . . n


1 + 2 + 3 + ... + n-1 + n = S
= number of white dots
n + n-1 + n-2 + ... + 2 + 1 = S
= number of yellow dots
(n+1) + (n+1) + (n+1) + ... + (n+1) + (n+1) = 2S
n (n+1) = 2S
There are n(n+1) dots in the grid
n n n n n n
n+1 n+1 n+1 n+1 n+1


1 + 2 + 3 + ... + n-1 + n = S
= number of white dots
n + n-1 + n-2 + ... + 2 + 1 = S
= number of yellow dots
(n+1) + (n+1) + (n+1) + ... + (n+1) + (n+1) = 2S
n (n+1) = 2S
$S = \frac{n(n+1)}{2}$
n n n n n n
n+1 n+1 n+1 n+1 n+1


Very convincing! The unary representation brings out the geometry of the problem and makes each step look very natural.
By the way, my name is Bonzo. And you are?


Odette.
Yes, Bonzo. Let's take it even further…

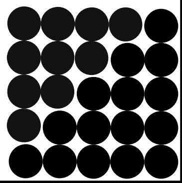

n^{th} Triangular Number
Δ_n = 1 + 2 + 3 + . . . + n-1 + n
= n(n+1)/2

n^{th} Square Number $= n^2$

$= \square_n$

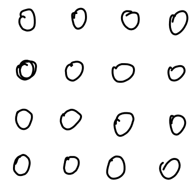
$n = \Delta_n + \Delta_{n-1}$

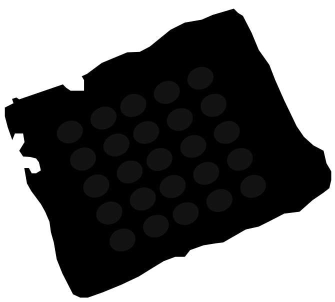
$= n^2$



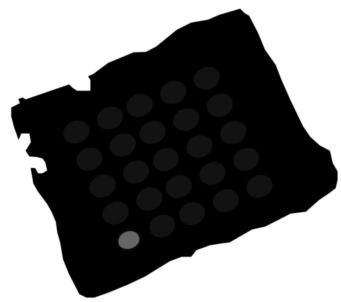
Breaking a square up in a new way.

$1 + 3 + 5 + 7 = 4^2$



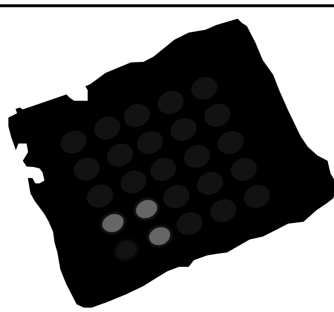


Breaking a square up in a new way.



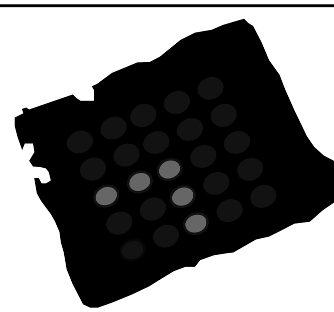
1

Breaking a square up in a new way.



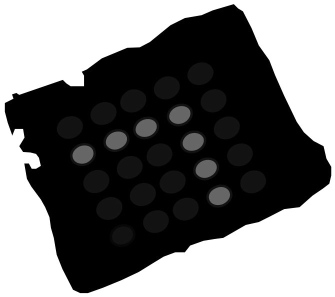
1 + 3

Breaking a square up in a new way.



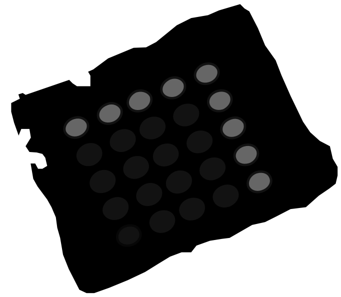
1 + 3 + 5

Breaking a square up in a new way.



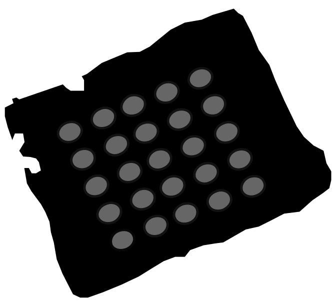
1 + 3 + 5 + 7

Breaking a square up in a new way.



1 + 3 + 5 + 7 + 9

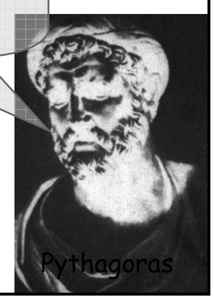
Breaking a square up in a new way.



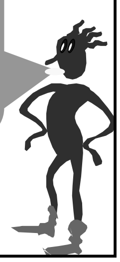
$1 + 3 + 5 + 7 + 9 = 5^2$

The sum of the first 5 odd numbers is 5 squared

The sum of the first n odd numbers is n squared.



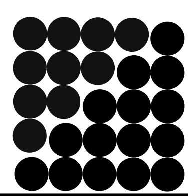
Here is an alternative dot proof of the same sum….



n^{th} Square Number

$n = \Delta_n + \Delta_{n-1}$

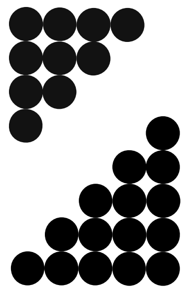
$= n^2$



n^{th} Square Number

n $= \Delta_n + \Delta_{n-1}$

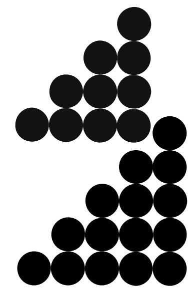
$= n^2$



n^{th} Square Number

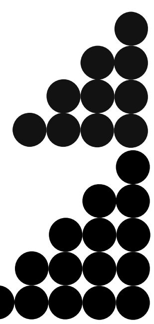
n $= \Delta_n + \Delta_{n-1}$

$= n^2$



Look at the columns!

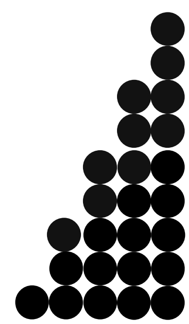
n $= \Delta_n + \Delta_{n-1}$



Look at the columns!

n $= \Delta_n + \Delta_{n-1}$

= Sum of first n odd numbers.



High School Notation

$\Delta_n + \Delta_{n-1} =$

$$
\begin{array}{r}
1 + 2 + 3 + 4 \ldots \\
+ \quad 1 + 2 + 3 + 4 + 5 \ldots \\
\hline
1 + 3 + 5 + 7 + 9 \ldots
\end{array}
$$

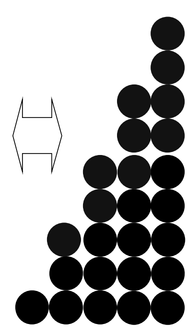
Sum of odd numbers

High School Notation

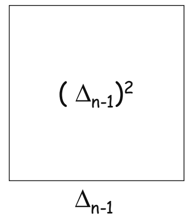
$\Delta_n + \Delta_{n-1} =$

$$
\begin{array}{r}
1 + 2 + 3 + 4 \ldots \\
+ \quad 1 + 2 + 3 + 4 + 5 \ldots \\
\hline
1 + 3 + 5 + 7 + 9 \ldots
\end{array}
$$

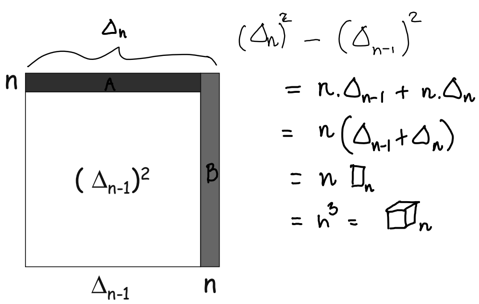
Sum of odd numbers



area of square is $(\Delta_{n-1})^2$



area of new square is $(\Delta_n)^2$

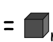


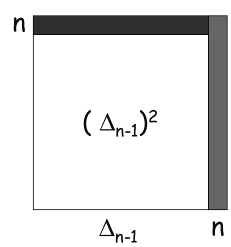
area of new square is $(\Delta_n)^2$

New shaded area = $n\Delta_n + n\Delta_{n-1}$

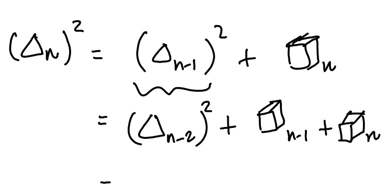
$= n\,(\Delta_n + \Delta_{n-1})$

$= n$ $_n$

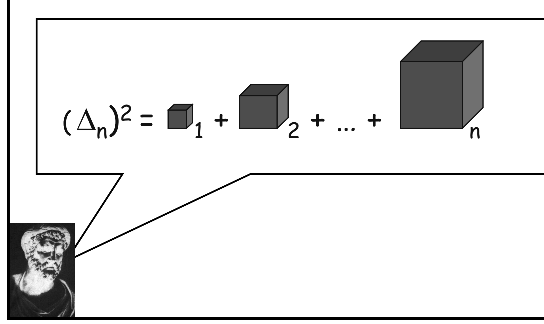
=  $_n$



$(\Delta_n)^2 = (\Delta_{n-1})^2 +$  $_n$



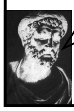
$(\Delta_n)^2 = (\Delta_{n-1})^2 +$  $_n$

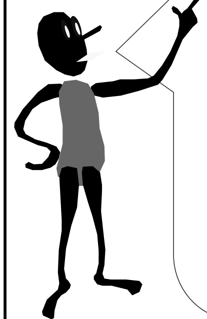


$(\Delta_n)^2 = (\Delta_{n-1})^2 +$  $_n$

$(\Delta_n)^2 = 1^3 + 2^3 + 3^3 + \ldots + n^3$

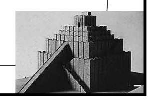
$= [\ n(n+1)/2\ ]^2$

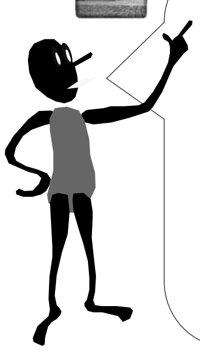




Can you find a formula for the sum of the first n squares?

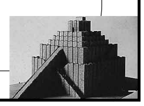
The Babylonians needed this sum to compute the number of blocks in their pyramids.

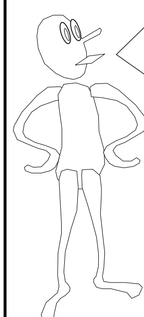




Can you find a formula for the sum of the first n squares?

$$\frac{n(n+1)(2n+1)}{6}$$

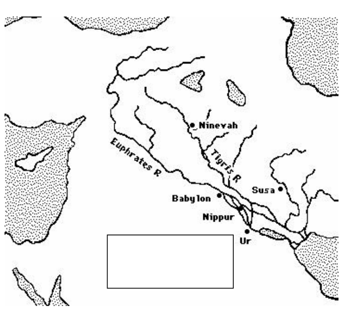




The ancients grappled with problems of abstraction in representation and reasoning.

Let's look back to the dawn of symbols…

Sumerians [modern Iraq]



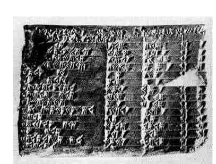
Sumerians [modern Iraq]

8000 BC Sumerian tokens use multiple symbols to represent numbers

3100 BC Develop Cuneiform writing

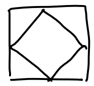
2000 BC Sumerian tablet demonstrates:
- base 10/60 notation (no zero)
- solving linear equations
- simple quadratic equations

Biblical timing: Abraham born in the Sumerian city of Ur

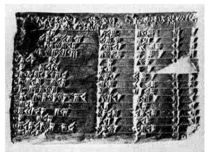


Babylonians absorb Sumerians

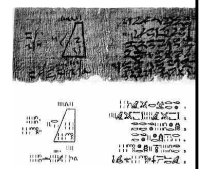
1900 BC Sumerian/Babylonian Tablet
- Sum of first n numbers
- Sum of first n squares
- “Pythagorean Theorem”
- “Pythagorean Triplets”, e.g., 3-4-5
- some bivariate equations



1600 BC Babylonian Tablet
- Take square roots
- Solve system of n linear equations



Egyptians



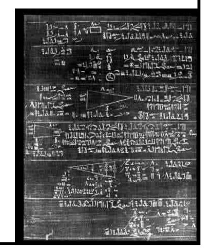
6000 BC Multiple symbols for numbers

3300 BC Developed Hieroglyphics

1850 BC Moscow Papyrus
- Volume of truncated pyramid

1650 BC Rhind Papyrus [Ahmes/Ahmose]
- Binary Multiplication/Division
- Sum of 1 to n
- Square roots
- Linear equations

Biblical timing: Joseph is Governor of Egypt.



Harrappans [Indus Valley Culture] Pakistan/India

3500 BC Perhaps the first writing system?!

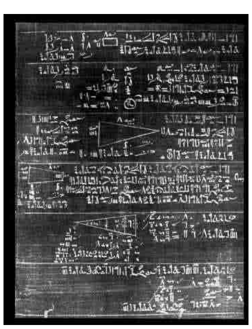
2000 BC Had a uniform decimal system of weights and measures

China

1200 BC Independent writing system
Surprisingly late.

1200 BC I Ching [Book of changes]
Binary system developed to do numerology.

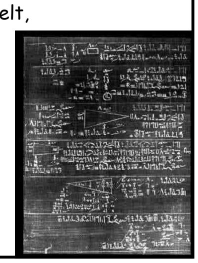
Rhind Papyrus: Scribe Ahmes was the Martin Gardener of his day!



Rhind Papyrus had 87 Problems.

A man has seven houses,
Each house contains seven cats,
Each cat has killed seven mice,
Each mouse had eaten seven ears of spelt,
Each ear had seven grains on it.
What is the total of all of these?

Sum of first five powers of 7



$$1 + X^1 + X^2 + X^3 + \ldots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

We'll use this fundamental sum again and again:

The Geometric Series

A Frequently Arising Calculation

$(X-1)\ (1 + X^1 + X^2 + X^3 + \ldots + X^{n-2} + X^{n-1})$

$= \quad X + X^2 + X^3 + \cdots + X^{n-1} + X^n$

$\quad -1 - X^1 - X^2 - X^3 - \cdots - X^{n-1}$

$\quad -1 + X^n$

A Frequently Arising Calculation

$(X-1)\ (1 + X^1 + X^2 + X^3 + \ldots + X^{n-2} + X^{n-1})$

$= \quad X^1 + X^2 + X^3 + \ldots + X^{n-1} + X^n$

$\quad - 1 - X^1 - X^2 - X^3 - \ldots - X^{n-2} - X^{n-1}$

$= \quad -1 \quad + X^n$

$= \quad X^n - 1$

$\Rightarrow 1 + X + X^2 + \cdots + X^{n-1} = \frac{X^n - 1}{X - 1}$ if $X \neq 1$

Action Shot: Mult by X is a SHIFT

$X\ (1 + X^1 + X^2 + X^3 + \ldots\ldots + X^{n-2} + X^{n-1})$

$= X\ (X^0 + X^1 + X^2 + X^3 + \ldots\ldots + X^{n-2} + X^{n-1})$

$= \quad X^1 + X^2 + X^3 + X^4 + \ldots\ldots + X^{n-1} + X^n$

The Geometric Series

$(X-1)\ (1 + X^1 + X^2 + X^3 + \ldots + X^{n-2} + X^{n-1}) = X^n - 1$

$$1 + X^1 + X^2 + X^3 + \ldots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

when $X \neq 1$

The Geometric Series for X=2

$1 + 2^1 + 2^2 + 2^3 + \ldots + 2^{n-1} = 2^n - 1$

$$1 + X^1 + X^2 + X^3 + \ldots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

(when $X \neq 1$)

The Geometric Series for X=3

$$1 + 3^1 + 3^2 + 3^3 + \dots + 3^{n-1} = (3^n - 1)/2$$

$$1 + X^1 + X^2 + X^3 + \dots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

(when $X \neq 1$)

The Geometric Series for X=½

$$1 + \tfrac{1}{2}^1 + \tfrac{1}{2}^2 + \tfrac{1}{2}^3 + \dots \tfrac{1}{2}^{n-1} = (\tfrac{1}{2}^n - 1)/(-\tfrac{1}{2})$$
$$= 2 - (\tfrac{1}{2})^{n-1}$$

$$1 + X^1 + X^2 + X^3 + \dots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

(when $X \neq 1$)

The Geometric Series

$$1 + ab + a^2b^2 + a^3b^3 + \dots + a^nb^n$$
$$= 1 + ab + (ab)^2 + (ab)^3 + \quad + (ab)^n$$
$$= \frac{(ab)^{n+1} - 1}{ab - 1} = \frac{a^{n+1}b^{n+1} - 1}{ab - 1}$$

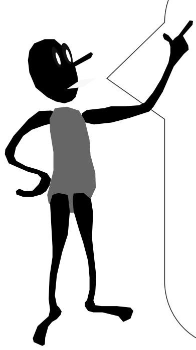
$$1 + X^1 + X^2 + X^3 + \dots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

(when $X \neq 1$)

The Geometric Series

$$1 + X^1 + X^2 + X^3 + \dots + X^{n-2} + X^{n-1} = \frac{X^n - 1}{X - 1}$$

(when $X \neq 1$)



Numbers and their properties can be represented as strings of symbols.

Strings Of Symbols.

We take the idea of symbol and sequence of symbols as primitive.

Let Σ be any fixed finite set of symbols.
Σ is called an alphabet, or a set of symbols.

Examples:
- Σ = {0,1,2,3,4}
- Σ = {a,b,c,d, …, z}
- Σ = all typewriter symbols.
- Σ = { a , b , c , d , …, z }

Strings over the alphabet Σ.

A string is a sequence of symbols from Σ.

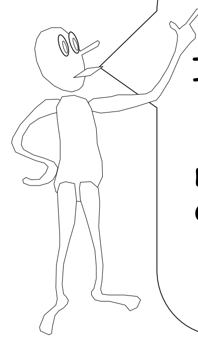
Let s and t be strings.
Then st denotes the concatenation of s and t
i.e., the string obtained by the string s followed by the string t.

ab bcd ⇒ abbcd

Now define Σ^+ by these inductive rules:

$x \in \Sigma \Rightarrow x \in \Sigma^+$

$s,t \in \Sigma^+ \Rightarrow st \in \Sigma^+$



Intuitively, Σ^+ is the set of all finite strings that we can make using (at least one) letters from Σ.

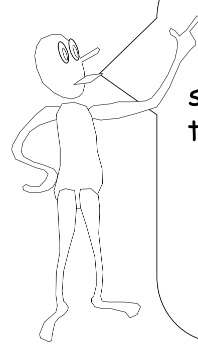
The set Σ^*

Define ε be the empty string.
I.e., XεY= XY for all strings X and Y.

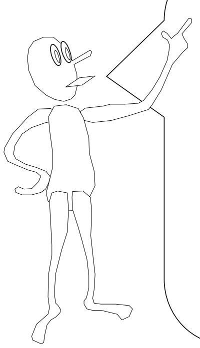
ε is also called the string of length 0.

Define $\Sigma^0 = \{\varepsilon\}$

Define $\Sigma^* = \Sigma^+ \cup \{\varepsilon\} = \Sigma^+ \cup \Sigma^0$

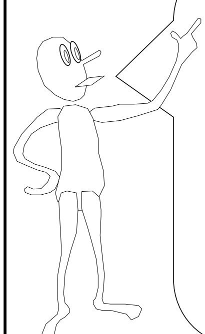


Intuitively, Σ^* is the set of all finite strings that we can make using letters from Σ, including the empty string.



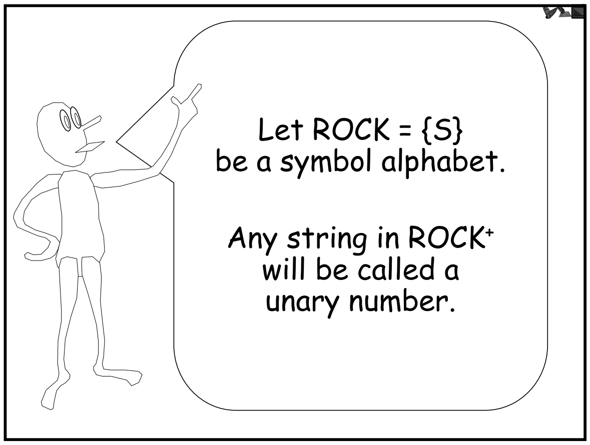
Let DIGITS = {0,1,2,3,4,5,6,7,8,9} be a symbol alphabet.

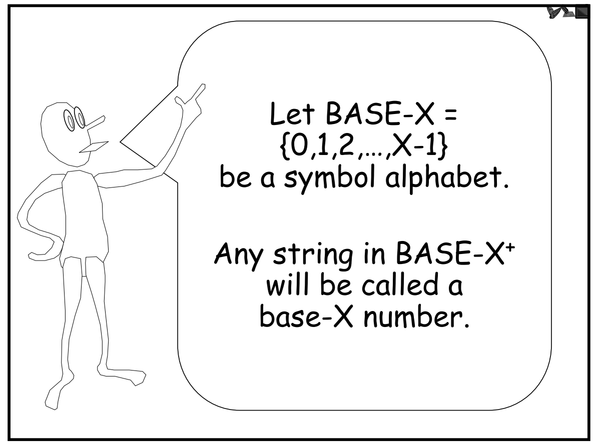
Any string in $DIGITS^+$ will be called a decimal number.

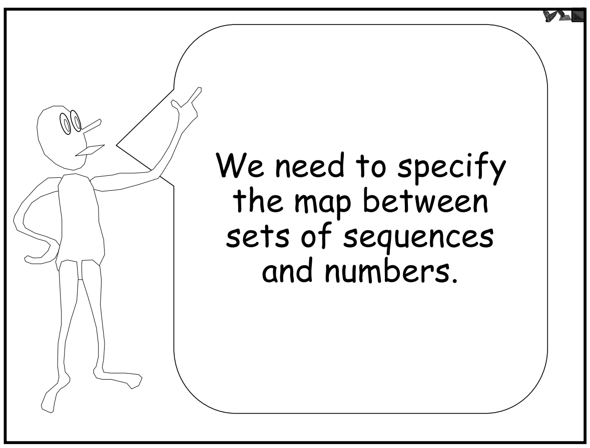


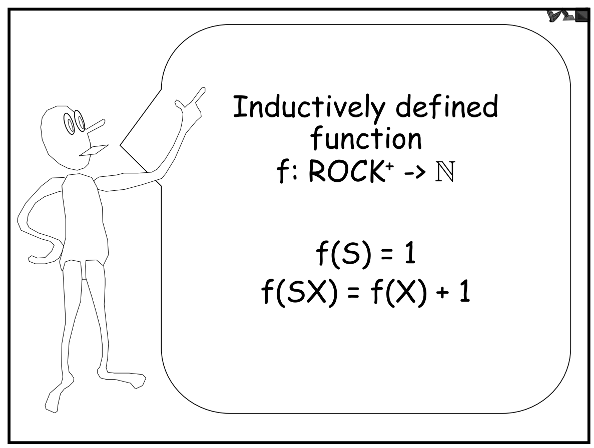
Let BITS = {0,1} be a symbol alphabet.

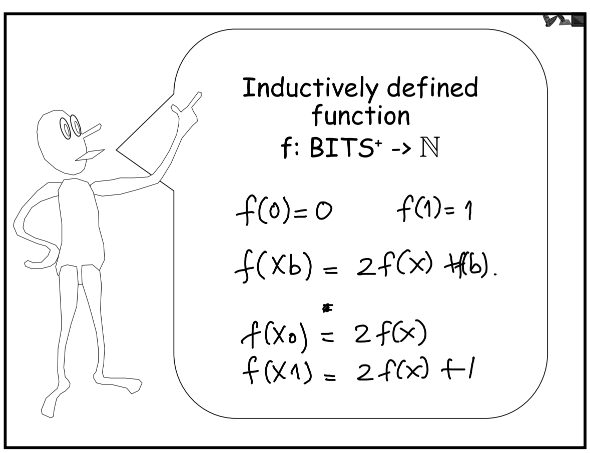
Any string in $BITS^+$ will be called a binary number.

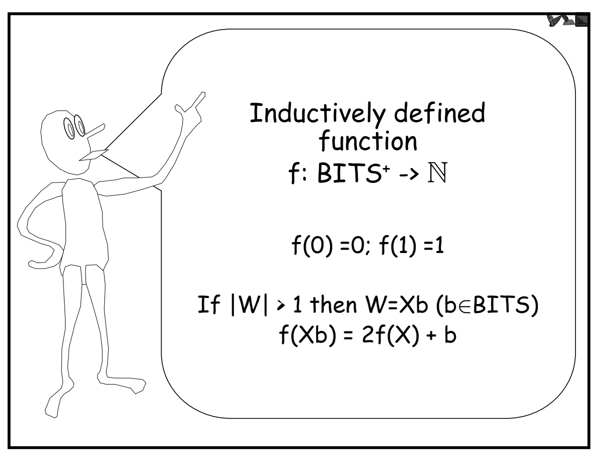

Let ROCK = {S}
be a symbol alphabet.
Any string in ROCK+
will be called a
unary number.

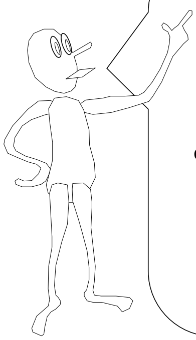

Let BASE-X =
{0,1,2,...,X-1}
be a symbol alphabet.
Any string in BASE-X+
will be called a
base-X number.


We need to specify
the map between
sets of sequences
and numbers.


Inductively defined
function
f: ROCK+ -> ℕ
f(S) = 1
f(SX) = f(X) + 1


Inductively defined
function
f: BITS+ -> ℕ
f(0)=0 f(1)=1
f(Xb) = 2f(X) +f(b).
f(X0) = 2f(X)
f(X1) = 2f(X) +1


Inductively defined
function
f: BITS+ -> ℕ
f(0) =0; f(1) =1
If |W| > 1 then W=Xb (b∈BITS)
f(Xb) = 2f(X) + b



Non-inductive representation of f:

$g(a_{n-1}\, a_{n-2} \dots a_0) =$
$a_{n-1} * 2^{n-1} + a_{n-2} * 2^{n-2} + \dots + a_0 2^0$

Theorem: f and g are identical

$f(0) = 0;\ f(1) = 1$
$f(Xb) = 2f(X) + b$ for $b \in BITS$

$g(a_{n-1}\, a_{n-2} \dots a_0) =$
$a_{n-1} * 2^{n-1} + a_{n-2} * 2^{n-2} + \dots + a_0 2^0$

Induction on the length of the string.

Basecase: length = 1

$\Rightarrow f(0) = 0 \qquad g(0) = 0$
$f(1) = 1 \qquad g(1) = 1$ ✓

Theorem: f and g are identical

$f(0) = 0;\ f(1) = 1$
$f(Xb) = 2f(X) + b$ for $b \in BITS$

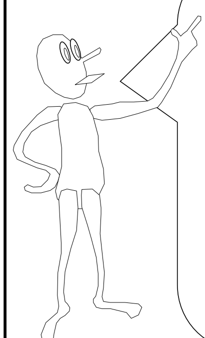
$g(a_{n-1}\, a_{n-2} \dots a_0) =$
$a_{n-1} * 2^{n-1} + a_{n-2} * 2^{n-2} + \dots + a_0 2^0$

I.H. $f = g$ for all strings of length $< n$.

Inductive Step:

Consider a number of length n. $a_{n-1} a_{n-2} \dots a_0$

$f(a_{n-1} \dots a_0) = 2f(a_{n-1} \dots a_1) + a_0$
$= 2g(a_{n-1} \dots a_1) + a_0$ (by I.H.)
$= 2(a_{n-1} 2^{n-2} + \dots + a_1 . 2^0) + a_0$
$= a_{n-1} 2^{n-1}$ [illegible] $\dots + a_0$

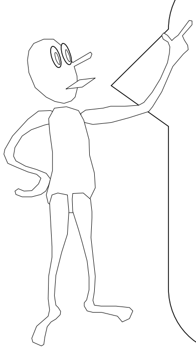


Two identical maps from sequences to numbers:

$f(0) = 0;\ f(1) = 1$
$f(Xb) = 2f(X) + b$

and

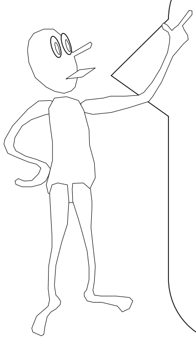
$f(a_{n-1}\, a_{n-2} \dots a_0) =$
$a_{n-1} * 2^{n-1} + a_{n-2} * 2^{n-2} + \dots + a_0 2^0$



The symbol a_0 is called the Least Significant Bit or the Parity Bit.

$a_0 = 0$
iff
$f(a_{n-1} a_{n-2} \dots a_0) =$
$a_{n-1} * 2^{n-1} + a_{n-2} * 2^{n-2} + \dots + a_0 2^0$
is an even number.

Theorem: Each natural number has a binary representation.



Theorem: Each natural has a binary representation.

Base Case: 0 and 1 do.

Induction hypothesis: Suppose all natural numbers less than n have a binary representation.

Induction Step: Note that n=2m+b for some m<n, with b=0 or 1. Represent n as the left-shifted sequence for m concatenated with the symbol for b.

No Leading Zero Binary (NLZB)

A binary string that is either 0 or 1,
Or has length > 1, and does not have a leading zero.

0, 1, 10, 100000001, 1110000 are in NLZB
01, 000001010111, 011111111 are not.

Theorem: Each natural number has a unique NLZBinary representation.

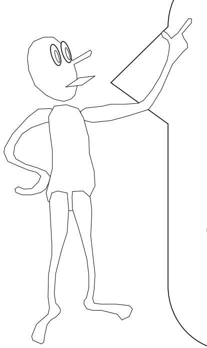
Base Case: 0 and 1 do.

Inductive Hypothesis: statement is true for all numbers < n

Then Inductive Step: Suppose n has 2 repr.

Wb
W'b' ⇒ b = b' and hence W = W'

Theorem: Each natural number has a unique NLZBinary representation.

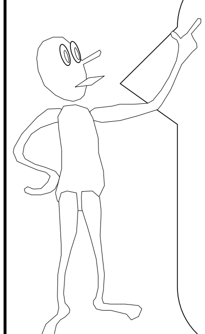


Theorem: Each natural has a unique NLZB representation.

Base Case: 0 and 1 do.

Induction Hypothesis: Every natural number less than n has a unique NLZB representation.

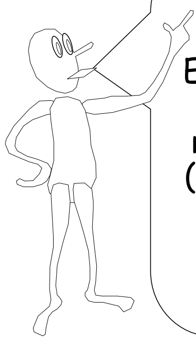
Induction Step: Suppose n=2m+b has 2 NLZB representations W and V. Their parity bit b must be identical. Hence, m also has two distinct NLZB representations, which contradicts the induction hypothesis. So n must have a unique representation.



Inductive definition is great for showing UNIQUE representation:

$f(Xb) = 2f(X) + b$

Let n be the smallest number reprinted by two different binary sequences. They must have the same parity bit, thus we can make a smaller number that has distinct representations.



Each natural number has a unique representation as a (No Leading Zeroes) Binary number!

BASE X representation

$S = a_{n-1}\, a_{n-2} \ldots a_1\, a_0$ represents the number:
$a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \ldots + a_0 X^0$

Base 2 [Binary Notation]
101 represents $1\,(2)^2 + 0\,(2^1) + 1\,(2^0)$
= ◯◯◯◯◯

Base 7
015 represents $0\,(7)^2 + 1\,(7^1) + 5\,(7^0)$
= ◯◯◯◯◯◯◯◯◯◯◯◯

Bases In Different Cultures

Sumerian-Babylonian: 10, 60, 360
Egyptians: 3, 7, 10, 60
Africans: 5, 10
French: 10, 20
English: 10, 12, 20

BASE X representation

$S = (a_{n-1}\, a_{n-2} \ldots a_1\, a_0)_X$ represents the number:
$a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \ldots + a_0 X^0$

Largest number representable in base-~~X~~ 2 with n "digits"

$= 1.2^{n-1} + 1.2^{n-2} + \ldots + 1$

$= 2^n - 1.$

00000 . ·0

11111 11

BASE X representation

$S = (a_{n-1}\, a_{n-2} \ldots a_1\, a_0)_X$ represents the number:
$a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \ldots + a_0 X^0$

Largest number representable in base-X with n "digits"

BASE X representation

$(111)_2$
$=(7)_{10}$

$S = (a_{n-1}\, a_{n-2} \ldots a_1\, a_0)_X$ represents the number:
$a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \ldots + a_0 X^0$

Largest number representable in base-X with n "digits"

$= (X\text{-}1\ X\text{-}1\ X\text{-}1\ X\text{-}1\ X\text{-}1 \ldots X\text{-}1)_X$

$= (X-1)(X^{n-1} + X^{n-2} + \ldots + X^0)$

$= (X^n - 1)$ $2^n - 1$

7 in 3bit
8 ⇒ 4bit

Fundamental Theorem For Binary:

Each of the numbers from 0 to 2^n-1 is uniquely represented by an n-bit number in binary.

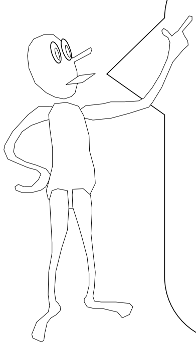
k uses $\lfloor \log_2 k \rfloor + 1$ digits in base 2.

$= \lceil \log_2 (k+1) \rceil$

Fundamental Theorem For Base-X:

Each of the numbers from 0 to X^n-1 is uniquely represented by an n-"digit" number in base-X.

k uses $\lfloor \log_X k \rfloor + 1$ digits in base-X.



n has length n in unary, but has length $\lfloor \log_2 n \rfloor + 1$ in binary.

Unary is exponentially longer than binary.

Other Representations: Egyptian Base 3

Conventional Base 3:
Each digit can be 0, 1, or 2

$(5)_{10} = (12)_3$

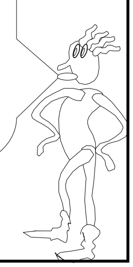
$= 3 \cdot 1 + 3^0 \cdot 2 = 5$

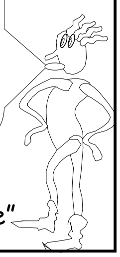
Here is a strange new one:
Egyptian Base 3 uses -1, 0, 1

Example: (1 -1 -1) = 9 - 3 - 1 = 5

We can prove a unique representation theorem

How could this be Egyptian? Historically, negative numbers first appear in the writings of the Hindu mathematician Brahmagupta (628 AD).





One weight for each power of 3.
Left = “negative”. Right = “positive”

Study Bee

Unary and Binary
- Triangular Numbers
- Little Gauss’s Proof
- Dot proofs

The Geometric Series
- $(1+x+x^2 + \dots + x^{n-1}) = (x^n - 1)/(x-1)$

Base-X representations
- unique binary representations
  - proof for no-leading zero binary
- k uses $\lfloor \log_2 k \rfloor + 1 = \lceil \log_2 (k+1) \rceil$ digits in base 2
  (unary is exponentially longer than binary)
- largest n-bit number in base X is $(X^n - 1)$