- More generally, we define a *ring* to be a non-empty set R having two binary operations (we'll think of these as addition and multiplication) which is an Abelian group under + (we'll denote the additive identity by *0*), and satisfies the following additional properties:

  (v) $a(bc) = (ab)c \ \ \forall\, a, b, c \in R$ i.e., associativity for multiplication

  (vi) $\exists\ 1 \in R$ such that $1a = a = a1 \ \ \forall\, a \in R$ i.e., a multiplicative identity*

  (vii) $a(b + c) = (ab) + (ac)$ and $(a + b)c = (ac) + (bc) \ \ \forall\, a, b, c \in R$ i.e., distributivity**

- Natural examples of rings are the ring of integers, a ring of polynomials in one variable, the ring $\mathbb{Z}_n$ of integers mod *n*, the Boolean ring of $\mathscr{P}(A)$ for some set A under + (symmetric difference) and ∩ (as multiplication), and the ring of 3 × 3 matrices of real numbers. Since + is always commutative in a ring, we say that a ring is *commutative* if its multiplication is commutative (notice that the matrix ring fails to be commutative).

- Notice also that there's no requirement for a ring to have multiplicative inverses. If a ring is commutative *and* has multiplicative inverses for everything other that *0*, then it's called a *field*. For example, the field of rational numbers, or of real numbers, or integers modulo a prime, or non-singular 3 × 3 matrices together with the zero matrix.

- The fact that the integers are a commutative ring, but fail to be a field, and that they're very useful(!!), is why attention is focussed on rings. In this course we'll mostly confine our attention to commutative rings because of the more immediate application to the integers.

---

* Some authors prefer not to include the requirement of a general ring having a multiplicative identity, preferring to talk about *rings* and *rings with identity* as very distinct animals. There are some benefits to this, however the community is divided on this, so for simplicity we'll go with the side that includes *1*.

** There are a bunch of useful properties common to all rings. For example, $0x = 0$ always, since $0x = (0 + 0)x = 0x + 0x$. Also $(-1)x = -x$ always, since $x + (-1)x = 1x + (-1)x = (1 + (-1))x = 0x = 0$. Similarly, $(-a)(-b) = ab$ always.

- Returning to where we left off two slides ago, we observed that in the field $\mathbb{Z}_p$ any non-zero element a has the property that $a^{p-1} = 1$. What about the commutative ring $\mathbb{Z}_n$ when n isn't a prime, for example when $n = p^r$?

- First some definitions. Let's define a *unit* in a commutative[*] ring to be any element $u \in \mathrm{R}$ such that $\exists\, v \in \mathrm{R}$ with $uv = 1$. The set of all units in a ring (clearly non=empty since 1 is a unit) is called the group of units of the ring R , and is often denoted $\mathrm{R}^\times$.

- We define $p \in \mathrm{R}$ to be a *prime* if it's not a unit and if $p \mid ab \implies p \mid a$ or $p \mid b$.[**]

- Similarly, we can define a *zero divisor* to be any non-zero element $r \in \mathrm{R}$ such that $\exists\, s \in \mathrm{R}$ with $s \neq 0$ and $rs = 0$. Notice that zero divisors can't have multiplicative inverses.[***] If $t \in \mathrm{R}$ is such that $t^k = 0$ for some $k \geq 1$, then t is said to be *nilpotent*.

- Notice that if a ring has no zero divisors[****], then even if it doesn't have multiplicative inverses, we can still solve equations like $ax = ab$ if $a \neq 0$, for then $0 = ax - ab = a(x - b)$, and the lack of zero divisors then implies that either $a = 0$ or $x - b = 0$, hence $x - b = 0$ and so $x = b$. So 'cancellation' is still possible even without inverses!

- So what does the group of units of $\mathrm{R} = \mathbb{Z}_n$ look like? We could list the elements, so if $n = p^r$:

$$\mathrm{R}^\times = \{ \cancel{0}, 1, 2, \ldots, p-1, \cancel{p}, p+1, \ldots, 2p-1, \cancel{2p}, 2p+1, \ldots, 3p-1, \cancel{3p}, 3p+1, \ldots\ldots, p^{r-1}-1, \cancel{p^{r-1}}, p^{r-1}+1, \ldots, p^r - 1 \}$$

- So there were a total of p^r elements in R, from which we've deleted p^{r-1} of them (since multiples of p can't have multiplicative inverses), leaving us with $|\mathrm{R}^\times| = p^r - p^{r-1}$.

---

* If defining this for a non-commutative ring, then we'd have to require both that $uv = 1$ and $vu = 1$.

** Recall that the notation $x \mid y$ means that x divides y exactly. Notice also that our definition specifically excludes 1 from being a prime.

*** Otherwise if $mk = 1$ and $rm = 0$, then $r = r1 = r(mk) = (rm)k = 0k = 0$.

**** Such a ring is called an *integral domain*.

- More generally, if n is some composite number, then $n = rm$, whereupon m is a zero divisor and so can't have a multiplicative inverse. So $R^{\times}$ comprises numbers between 1 and n which are co-prime to n. This quantity is a function, namely $\varphi(n)$, and is called the *Euler phi function*.

- Notice that $a^{\varphi(n)} = 1$ for any non zero-divisor $a \in R^{\times}$ when $R = \mathbb{Z}_n$.

- How can we calculate $\varphi(n)$? We've seen that $\varphi(n) = p - 1$ if $n = p$ (a prime), and also that $\varphi(n) = p^r - p^{r-1}$ if $n = p^r$. Indeed, it's not hard to show* that if a and b are co-prime, then $\varphi(ab) = \varphi(a)\,\varphi(b)$, hence for n any composite number, by factorising n completely we can get $\varphi(n) = \varphi(p^r \ldots q^z) = (p^r - p^{r-1}) \ldots (q^z - q^{z-1}) = n(1 - p^{-1}) \ldots (1 - q^{-1})$.

  - Let's play with what we have for a bit. We'll calculate 2^{35} (mod 7). Since $35 = (6)(5) + 5$ and $\varphi(7) = 6$, we can see that $2^{35} = (2^6)^5 \times 2^5 = 1 \times 32 = 4$ (mod 7).

  - Similarly, we can calculate $11^{81050696835}$ (mod 1176). Firstly we factorise $1176 = (2^3)(3)(7^2)$, so then $\varphi(1176) = (8 - 4)(3 - 1)(49 - 7) = 336$, and then compute 81050696835 (mod 336), namely 3. Hence $11^{81050696835} = (11^{336})^{241222312} \times 11^3 = 1331$ (mod 1176) = 155 (mod 1176).

- In order to be able to handle multiple *simultaneous* modular equations, we'll need to be able to relate multiple rings. Essentially, each ring will 'define' the effect of each modular factor, but there's a natural isomorphism which will make life much easier. Let's start by supposing that we have two rings R and S , then we can make the set $R = R \times S$ into a ring in a natural way by defining $(r, s) + (r', s') = (r + r', s + s')$ and $(r, s)(r', s') = (rr', ss')$.

---

* you can check that this is true simply by comparing the number of zero divisors of *a, b,* and *ab*. We can also get this as a corollary of the Chinese remainder theorem we'll prove shortly.

- Define the subset $A \subseteq R$ to be *absorptive** if it's a subgroup of R under $+$ and that $rA \subseteq A$ and $Ar \subseteq A$ ** for all $r \in R$ (hence the use of the word 'absorptive'). As a natural example of such a set in the ring $\mathbb{Z}$ consider $A = n\mathbb{Z}$, since for any $r \in R$, any $t \in rA$ is automatically a multiple of n and so lives in A.

- It's not hard to show that if $f : R \longrightarrow S$ is a ring homomorphism, then $\ker f$ is an absorptive set (the kernel for rings will be the set of all elements of R which get mapped to 0_S). Notice that if A is absorptive and $1 \in A$, then $A = R$, so typically we're more interested in the cases where A doesn't contain 1.

- We define $A + B = \{ a + b \mid a \in A,\ b \in B \}$ and $AB = \{ \Sigma\, a_i b_i \mid a_i \in A,\ b_i \in B \}$, where the summation in AB is only ever allowed to be over finitely many things. It's easy to show that both of these are absorptive if A and B are absorptive.

- If A and B are absorptive in R, then we can define $f : R \longrightarrow R/A \times R/B$ by $f(r) = (r + A, r + B)$. Then f is a (ring) homomorphism,*** with $\ker f = A \cap B$.

- If moreover $A + B = R$, then f is onto and $A \cap B = AB$,**** hence R/AB is isomorphic to $R/A \times R/B$. In the context of $R = \mathbb{Z}$ and $A = m\mathbb{Z}$ and $B = n\mathbb{Z}$, then if m and n are coprime, then $AB = mn\mathbb{Z}$. However, why is it obvious that $A + B = R$ in this case? It's all down to the fact that the gcd (greatest common divisor) of m and n is 1.

---

* This is *not* standard notation. The standard language is to call these sets *ideals*. However, the word 'absorptive' seems to characterise how they behave, so for now we'll use this more descriptive language.

** Notice that in fact, $rA = A = Ar$. That's easy to see once you spot that $1 \in R$ means that $1A \subseteq A$.

*** Showing $f(r + s) = f(r) + f(s)$ is straightforward. However, the absorptive nature of A and B is needed to show that $f(rs) = f(r)\, f(s)$ since, for example, $(r + A)(s + B) = rs + rA + Ar + AA = rs + A + A + A = (rs) + A$

**** Since A is absorptive, $a_i b_i \in A$, and since B is absorptive, $a_i b_i \in B$, and since absorptive sets are subgroups under +, $\Sigma\, a_i b_i \in A \cap B$ □ $AB \subseteq A \cap B$. Now if $A + B = R$ then $1 \in A + B$ so $1 = a + b$ for some $a \in A$ and $b \in B$, but then $c \in A \cap B$ □ $c = c1 = c(a + b) = ca + cb \in AB$. Hence $A \cap B \subseteq AB$.

- So we need an effective way to compute the gcd of a pair of numbers. Claim: if $d = \gcd(m,n)$ then the equation $mx + ny = kd$ has solutions $x, y \in \mathbb{Z} \ \ \forall k \in \mathbb{Z}$.

- We'll approach this by actually *constructing* a solution to the equation $13155x + 2367y = 3$ using the *Euclidean subtraction algorithm*.

  Divide *2367* into *13155* to get $13155 = 5 \times 2367 + 1320$
  Divide *1320* into *2367* to get $2367 = 1 \times 1320 + 1047$
  Divide *1047* into *1320* to get $1320 = 1 \times 1047 + 273$
  Divide *273* into *1047* to get $1047 = 3 \times 273 + 228$
  Divide *228* into *273* to get $273 = 1 \times 228 + 45$
  Divide *45* into *228* to get $228 = 5 \times 45 + 3$
  Divide *3* into *45* to get $45 = 15 \times 3 + 0$ STOP .
  Hence this last 'dividing value' is the gcd of *2367* and *13155* .

Hence if we let *units = 3,* then working from the bottom up,
45 = 15 units
228 = 5(45) + 3 = 5(15 units) + 1 unit = 76 units
273 = 1(228) + 45 = 1(76) + 15 units = 91 units
1047 = 3(273) + 228 = 3(91) + 76 units = 349 units
1320 = 1(1047) + 273 = 1(349) + 91 units = 440 units
2367 = 1(1320) + 1047 = 1(440) + 349 units = 789 units
13155 = 5(2367) + 1320 = 5(789) + 440 units = 4385 units

We can also use this information to get a solution to the equation by focussing on the remainders, letting $a = 13155$ and $b = 2367$, via

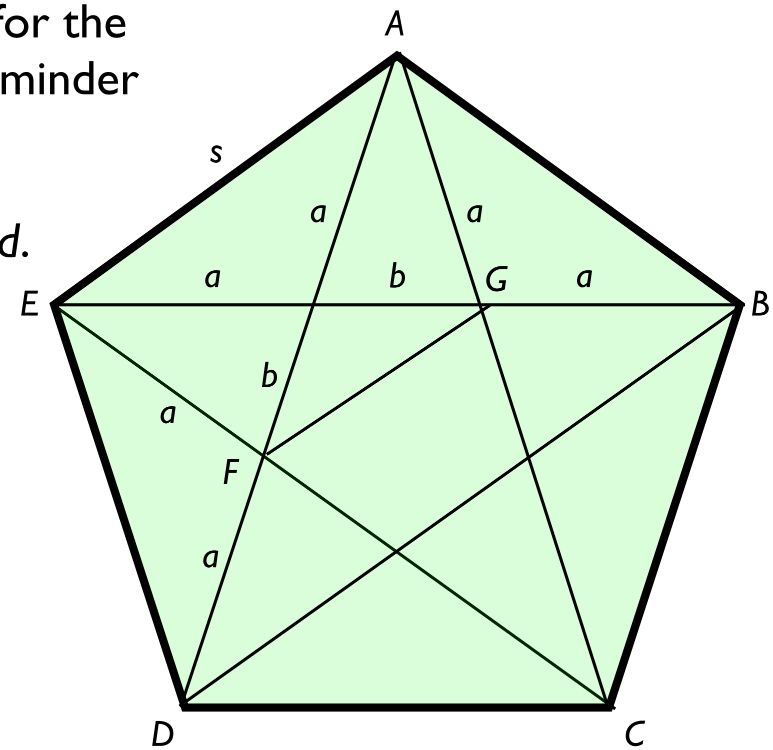
$1320 = a - 5b$
$1047 = b - 1320 = b - (a - 5b) = -a + 6b$
$273 = 1320 - 1047 = (a - 5b) - (-a + 6b) = 2a - 11b$
$228 = 1047 - 3(273) = (-a + 6b) - 3(2a - 11b) = -7a + 39b$
$45 = 273 - 228 = (2a - 11b) - (-7a + 39b) = 9a - 50b$
$3 = 228 - 5(45) = (-7a + 39b) - 5(9a - 50b) = -52a + 289b$

- This is actually a process to find a largest *common* measurement unit for two lengths; at least, that was how it was formulated in ancient Greek times. It was used to get approximations for the ratios of a side to the diagonal of a square, for a side to the diagonal of a regular pentagon, and for the ratio of the circumference to the diameter of a circle, and is strongly related to the construction of *continued fractions*.

$$\frac{13155}{2367} = 5 + \frac{1320}{2367} = 5 + \frac{1}{1 + \frac{1047}{1320}} = 5 + \frac{1}{1 + \frac{1}{1 + \frac{273}{1047}}} = \ldots = 5 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{5 + \frac{1}{15}}}}}}$$

- Notice that if we take successive truncations of the previous continued fraction, we get a sequence of rational approximations to the value of $r = 13155 \div 2367$, and that they alternate around above and below, namely: $r \approx 5$, 6, $5+ 1/2$, $5+ 4/7$, $5+ 5/9$, $5+ 29/52$, $5+ 440/789$, or in decimal form: $r \approx \uparrow 5$, $\downarrow 6$, $\uparrow 5.5$, $\downarrow 5.57143$, $\uparrow 5.55556$, $\downarrow 5.55769231$, $=5.55766793$.

- We'll apply this to find an sequence of approximations for the diameter : side ratio for a regular pentagon. It'll be a reminder of euclidean geometry ....



- We'll label the length of a side by s and a diagonal by d.

  So in our picture, $d = 2a + b$.

  Since AB is parallel to EC, AEF is isoceles, and $AF = s$.

  So $d = s + a$ and $s = a + b$

  Since EA is parallel to GF and BD, EFG is isoceles, and $GF = a$.

  But GF is the diagonal of another regular pentagon with side b.

  So $d : s$ is the same ratio as $b : a$.

  So in the spirit of the euclidean subtraction algorithm ...

  $d = 1s + a$, $s = 1a + b$, $b = 1s_1 + a_1$, $s_1 = 1a_1 + b_1$, $b_1 = 1s_2 + a_2$, $s_2 = 1a_2 + b_2$, $b_2 = 1s_3 + a_3$, ...

  where the b_n, s_n, and a_n are repeating the same pattern of sides and diagonals within ever smaller pentagons.

  This gives a continued fraction expansion: $d / s = 1 + 1/ (1 + 1/ (1 + 1/ (1 + 1/ (1 + \ldots\ldots))))$ which gives successive rational approximations: 1, $1+ 1/2$, $1+ 2/3$, $1+ 3/5$, $1+ 5/8$, $1+ 8/13$, $1+ 13/21$, $1+ 21/34$, ...

  Not only does this method give a surprisingly fast rational approximation, but it also amusingly uses the Fibonacci sequence!

- Applied to estimating π this process yields: 3, $22/7$, $333/106$, $355/113$, $103993/33102$, ...

- This approach can be generalised. Suppose a ring R can admit a function $f : R^* \longrightarrow \mathbb{N}$ such that $a, b \in R^* \implies \exists\, q, r \in R$ with $a = bq + r$ and $f(r) < f(b)$ or $r = 0$. Certainly the ring of integers satisfies this (with for example, $f(n) = |n|$ being the function). Can we do this for $\mathbb{Z}_n$? Any ring which has this property is said to be a *Euclidean domain*.

- Notice that we could extend this to the ring of polynomials having integer (or rational, or real, or complex) coefficients by defining f to yield the degree of the polynomial. The results of applying the Euclidean subtraction algorithm to such polynomials will of course be different depending on the range of coefficients available.

- Before our detour on finding gcds, we raised the question of solving multiple simultaneous modular equations, and observed that for A and B absorptive in R with A + B = R, then R/AB is isomorphic to R/A × R/B , and so if m and n are coprime, then $\mathbb{Z}_{mn}$ is isomorphic to $\mathbb{Z}_m \times \mathbb{Z}_n$. This is known as the *Chinese remainder theorem*.

- A consequence of this is that if we're given a pair of equations $x = a \bmod m$, $x = b \bmod n$ with m, n coprime, then $(a , b) \in \mathbb{Z}_m \times \mathbb{Z}_n$ corresponds to a unique value in $\mathbb{Z}_{mn}$, hence we can solve in the traditional simultaneous equation style, substituting successively.* We can even do this for many equations, as long as the moduli are pairwise coprime.

Suppose $x = 3 \bmod 11$, $x = 6 \bmod 8$, $x = -1 \bmod 15$.

The first equation gives us $x = 3 + 11a$, so from the second equation we get $x = 6 + 8b = 3 + 11a$.

This gives $3 = 11a - 8b \implies a = 1 = b \implies x = 14 + 88t$ as a general solution for the first two equations.

The third equation gives us $x = -1 + 15c = 14 + 88t \implies 15 = 15c - 88t \implies c = 1,\ t = 0$

$\implies x = 14 + 1320s$ is the general solution for the three equations.

* This approach actually feels a bit like using the Euclidean algorithm -- we'll work through a more efficient way to formulate this in the homework.