Atari 2600 Programming for Newbies

By Andrew Davie

# Session 1: Start Here

So, you want to program the Atari 2600 and don't know where to start?

Welcome to the first installment of "000001010 00101000 00000000 1100101"—which at first glance is a rather odd name for a programming tutorial—but on closer examination is appropriate, as it is closely involved with what it's like to program the Atari 2600. The string of 0's and 1's is actually a binary representation of "2600 101".

I'm Andrew Davie, and I've been developing games for various computers and consoles since the late 1970s. Really! What I plan to do with this tutorial is introduce you to the arcane world of programming the '2600, and slowly build up your skill base so that you can start to develop your own games. We'll take this in slow easy stages.

Developing for the Atari 2600 is much simpler today than it was when the machine was a force in the marketplace (back in the 1980s). We have a helpful online community of dedicated programmers, readily available documentation, tools, and sample code—and online forums where we can pose questions and get almost instant feedback and answers. So don't be scared—with a bit of effort, anyone can do this!

It is this online community which makes developing for the machine 'fun'—though I use that in the broadest sense of the word. My 'fun' may be another man's 'torture'. For programming this machine is tricky at best—and not for the feint of heart. But the rewards are great—making this simple hardware do anything at all is quite an achievement—and making it do something new and interesting gives one a warm fuzzy feeling inside.

So, let's get right into it … here's your first installment of "2600 101". We're going to assume that you know how to program \*something\*, but not much more than that. We'll walk through binary arithmetic, hexadecimal, machine architecture, assemblers, graphics, and whatever else gets in our way. And we'll probably divert on tangential issues here and there. But hopefully we'll come out of it with a greater understanding of this little machine, and appreciation for the work of those brilliant programmers who have developed the classics for this system.

## The Basics

A game on the '2600 comes in the form of a cartridge (or 'tape') which is plugged into the console itself. This cartridge consists of a circuit board containing a ROM (or EPROM) which is basically just a silicon chip containing a program and graphics for displaying the game on your TV set. This program (and graphics) are really just a lot of numbers stored on the ROM which are interpreted by the CPU (the processor) inside your '2600 just like a program on any other computer. What makes the '2600 special is … nothing. It's a computer, just like any other!

A computer typically consists of a CPU, memory, and some input/output (I/O) systems. The '2600 has a CPU (a 6507), memory (RAM for the program's calculations, ROM to hold the program and graphics), and I/O systems (joystick and paddles for input, and output to your TV).

### The CPU

The CPU of the '2600 is a variant of a processor used in computers such as the Apple II, the Nintendo NES, the Super Nintendo, and Atari home computers (and others). It's used in all these machines because it is cheap to manufacture, it's simple to program, but also effective—the famous '6502'. In this course we will learn how to program the 6502 microprocessor … but don't panic, we'll take that in easy stages (and besides, it's not as hard as it looks).

The '2600 actually uses a 6507 microprocessor—but this is really just a 6502 dressed in sheep's clothing. The 6507 is able to address less memory than the 6502 but is in all other respects the same. I refer to the '2600 CPU as a 6502 purely as a matter of convenience.

### Memory

Memory is severely restricted on the '2600. When the machine was developed, memory (both ROM and RAM) were very expensive, so we don't have much of either. In fact, there's only 128 BYTES of RAM (and we can't even use all of that!)—and typically (depending on the capabilities of the cartridge we're going to be using for our final game) only about 4K of ROM. So, then, here's our first introduction to the 'limitations' of the machine. We may all have great ideas for '2600 games, but we must keep in mind the limited amount of RAM and ROM!

[If you'd like to create and sell games with a lot more ROM and RAM, check out the Melody boards at AtariAge. And be sure to pick up a Harmony Cartridge for testing your games on a real Atari 2600.]

### Input/Output

Input to the '2600 is through interaction by the users with joystick and paddle controllers, and various switches and buttons on the console itself. There are also additional control devices such as keypads—but we won't delve much into those. Output is invariably through a television picture with sound (the game that we see on our TV).

So, there's not really much to it so far—we have a microprocessor running a program from ROM, using RAM, as required, for the storage of data—and the output of our program being displayed on a TV set. What could be simpler?

## The Development Process

Developing a game for the '2600 is an iterative process involving editing source code, assembling the code, and testing the resulting binary (usually with an emulator). Our first step is to gather together the tools necessary to perform these tasks.

### Source Code

'Source code' is simply one or more text files (created by the programmer and/or tools) containing a list of instructions (and 'encoded' graphics) which make up a game. These data are converted by the assembler into a binary which is the actual data placed on a ROM in a cartridge, and is run by the '2600 itself.

### Text Editor

To edit your source code, you need a text editor—and here the choice is entirely up to you. I use Microsoft Developer Studio myself, as I like its features—but any text editor is fine. Packages integrating the development process (edit/assemble/test) into your text editor are available, and this integration makes the process much quicker and easier (for example, Developer-Studio integration allows a double-click on an error line reported by the assembler, and the editor will position you on the very line in the source code causing the error).

### Assembler

To convert your source code into a binary form, we use an 'assembler'. An assembler is a program which converts assembly language into binary format (and in particular, since the '2600 uses a 6502-variant processor, we need an assembler that knows how to convert 6502 assembly code into binary). Pretty much all '2600 development these days is done using the excellent cross-platformAvailable for multiple machines such as Mac, Linux, Windows, etc. assembler 'DASM' which was written by Matt Dillon in about 1988.

DASM is available online. It would be a good idea to go there now and get a copy of DASM, and the associated support-files for '2600 development. In this course, we will be using DASM exclusively. We'll learn how to setup and use DASM shortly. [DASM is also included with batari Basic, so you could get it there if you don't trust other sources.]

### Emulator

Development of a game in the 1980s consisted of creating a binary imageWrite source code, assemble into binary. and then physically 'burning' the binary onto an EPROM, putting that EPROM onto a cartridge and plugging it into a '2600. This was an inherently slow process (trust me, I did this for NES development!) and it sometimes took 15 minutes just to see a change!

Nowadays, we are able to see changes to code almost immediately because of the availability of good emulators. An emulator is a program which pretends to be another machine/program. For example, a '2600 emulator is able to 'run' binary ROM images and display the results just as if you'd actually plugged a cartridge containing a ROM with that binary into an actual '2600 console. Today's '2600 emulators are very good indeed.

So, instead of actually burning a ROM, we're just going to pretend we've burned one—and look at the results by running this pretend-ROM on an emulator. And if there's a problem, we go back and edit our source code, assemble it to a binary, and run the binary on the emulator again. That's our iterative development process in action.

There are quite a few '2600 emulators available, but two of note are:

* Stella
* z26

Stella is your best choice if you're programming on non-Windows platform. I use Z26 for Windows development, as it is quite fast and appears to be very accurate. Either of these emulators is fine, and it's handy to be able to cross-check results on either.

We'll learn how to use these emulators later—but right now let's continue with the gathering of things we need…

Now that we have an editor, an assembler, and an emulator—the next important things are documentation and sources for information. There are many places on the 'net where you can find information for programming '2600, but perhaps the most important are:

* The Stella List
* AtariAge

## Documentation

And finally, documentation. A copy of the technical specifications of the '2600 hardware (the Stella Programmer's Guide) is essential…

* Stella Programmer's Guide

## Summary

OK, that's all we need. Here's a summary of what you should have…

* Text editor of your choice.
* DASM assembler and '2600 support files.
* Emulator (Z26 or Stella)
* Stella Programmer's Guide
* Bookmarks to AtariAge and the #Stella mailing list.

That's it for this session. Have a read of the Stella Programmer's Guide (don't worry about understanding it yet), and try installing your emulator (and play a few games for 'research' purposes). Next time we will make sure that our development environment is setup correctly, and start to discuss the principles of programming a '2600 game.

P.S. I can't promise to complete this 'course'—but hopefully what I do write will be interesting and helpful.

# Session 2: Television Display Basics

Hopefully you've been through the first part and have your editor, assembler, emulator and documentation ready to go. What we're going to look at now is a basic overview of how a television works, and why this is absolutely necessary prerequisite knowledge for the '2600 programmer. We're not going to cover a lot of '2600 specific stuff this time, but this is most definitely stuff you NEED TO KNOW!

Television has been around longer than you probably realize. Early mechanical television pictures were successfully broadcast in the '20s and '30s (yes, really!—see tvdawn.com). The mechanical 'scanning' technology utilized in these early television systems are no doubt the predecessors to the 'scanning' employed in our modern televisions.

A television doesn't display a continuous moving image. In fact, television displays static (non-moving) images in rapid succession—changing between images so quickly that the human eye perceives any movement as continuous. And even those static images aren't what they seem—they are really composed of lots of separate lines, each drawn one after the other by your TV, in rapid succession. So quick, in fact, that hundreds of them are drawn every image, and many images are drawn every second. In fact, the actual numbers are very important, so we'll have a look at those right now.

## NTSC, PAL, and SECAM

The Atari 2600 console was released in many different countries around the world. Not all of these countries use the same television 'system'—in fact there are three variations of TV systems (and there are three totally different variations of Atari 2600 hardware to support these systems). These systems are called NTSC, PAL, and SECAM. NTSC is used for the USA and Japan, PAL for many European countries, and Australia, and SECAM is used in France, some ex-French colonies (for example, Vietnam), and Russia. SECAM is very similar to PAL (625/50Hz), but I won't spend much time talking about it, as Atari SECAM units are incredibly rare, and little if any development is done for that format anyway. Interestingly, the differences in requirements for displaying a valid TV image for these systems leads to the incompatibility between cartridges made for NTSC, PAL and SECAM Atari units. We'll understand why, shortly!

### Images Per Second (Frequency)

A television signal contains either 60 images per second (on NTSC systems) or 50 images per second (on PAL systems). This is closely tied to the frequency of mains AC power in the countries which use these systems—and this is probably for historical reasons. In any case, it's important to understand that there are differences. Furthermore, NTSC images are 525 scanlines deep, and PAL images are 625 scanlines deep. From this, it follows that PAL images have more detail—but are displayed less frequently—or alternatively, NTSC images have less detail but are displayed more often. In practice, TV looks pretty much the same in both systems.

But from the '2600 point of view, the difference in frequency (50Hz vs. 60Hz) and resolution (625 scanlines vs. 525 scanlines) is important—very important—because it is the PROGRAMMER who has to control the data going to the TV. It is not done by the '2600 (!!)—the '2600 only generates a signal for a single scanline.

This is completely at odds with how all other consoles work, and what makes programming the '2600 so much 'fun'. Not only does the programmer have to worry about game mechanics—but he or she also has to worry about what the TV is doing (for example, what scanline it is drawing, and when it needs to start a new image, etc.).

Let's have a look at how a single image is drawn by a TV…

### Scanline

A television is a pretty amazing piece of 1930's technology. It forms the images we see by shining an electron beam (or 3, for color TVs) onto a phosphor coating on the front of the picture tube. When the beam strikes the phosphor, the phosphor starts to glow—and that glow slowly decreases in brightness until the phosphor is next hit by the electron beam. The TV 'sweeps' the electron beam across the screen to form 'scanlines'—at the same time as it sweeps, adjusting the intensity of the beam, so the phosphor it strikes glow brightly or dimly. When the beam gets to the end of a scanline, it is turned off, and the deflection circuitry (which controls the beam) is adjusted so that the beam will next start a little bit down, and at the start (far left-hand-side) of the next scanline. And it will then turn on, and sweep left-to-right to draw the next scanline. When the last scanline is drawn, the electron beam is turned off, and the deflection circuitry is reset so that the beam's position will next be at the top left of the TV screen—ready to draw the first scanline of the next frame.

This 'turning-off' and repositioning process—at the end of a scanline, and at the end of an image—is not instantaneous—it takes a certain amount of time for the electronics to do this repositioning, and we'll understand this when we come to talk about the horizontal blank (when the beam is resetting to the left of the next scanline) and the vertical blank (when the beam is resetting to the top left scanline on the screen). I'll leave that for a later session, but when we do come to it, you'll understand what the TV is doing at these points.

A fairly complex—but nonetheless simple-to-understand analog signal controls the sweeping of the electron beam across the face of the TV. First it tells the TV to do the repositioning to the start of the top left line of the screen, then it includes color and intensity information for the electron beam as it sweeps across that line, then it tells the TV to reposition to the start of the next scanline, etc., right down to the last scanline on the screen. Then it starts again with another reposition to the start… That's pretty much all we need to know about how that works.

The Atari 2600 sends the TV the "color and intensity information for the electron beam as it sweeps across that line", and a signal for the start of each new line. The '2600 programmer needs to feed the TV the signal to start the image frame.

#### Interlacing

A little side-track, here. Although I stated that the vertical resolution of a TV image is 625 lines (PAL) and 525 lines (NTSC), television employs another 'trick' called interlacing. Interlacing involves building up an image out of two separate 'frames'—each frame being either the odd scanlines, or the even scanlines of that image. Each frame is displayed every 1/30th of a second (30Hz) for NTSC, or every 1/25th of a second (25Hz) for PAL. By offsetting the vertical position of the start of the first scanline by half a scanline, and due to the persistence of the phosphor coating on the TV, the eye/brain combines these frames displaying alternate lines into a single image of greater vertical resolution than each frame. It's tricky and messy, but a glorious 'hack' solution to the problem of lack of bandwidth in a TV signal.

The upshot of this is that a single FRAME of a TV image is actually only half of the vertical resolution of the image. Thus, a NTSC frame is 525/2 = 262.5 lines deep, and a PAL frame is 625/2 = 312.5 lines deep. The extra .5 of a line is used to indicate to the TV if a frame is the first (even lines) or second (odd lines) of an image. An aside: about a year ago, the #stella community discussed this very aspect of TV images, and if it would be possible for the Atari to exploit this to generate a fully interlaced TV frame—and, in fact, it is possible. So some 25 years after the machine was first released, some clever programmers discovered how to double the resolution of the graphics.

Back to basics, though. We just worked out that a single frame on a TV is 262.5 (NTSC) and 312.5 (PAL) lines deep. And that extra .5 scanline was used to tell the TV if the frame was odd or even. So the actual depth of a single frame is 262 (NTSC) and 312 (PAL) lines. Now, if TV's aren't told that a frame is odd, they don't offset the first scanline by half a scanline's depth—and so, scanlines on successive frames are exactly aligned. We have a non-interlaced image, displayed at 60Hz (NTSC) or 50Hz (PAL). And this is the 'standard' format of an Atari 2600 frame sent to a TV.

In summary, an Atari 2600 frame consists of 262 scanlines (NTSC) or 312 scanlines (PAL), sent at 60Hz (NTSC) or 50Hz (PAL) frequency. It is the job of the '2600 programmer to make sure that the correct number of scanlines are sent to the TV at the right time, with the right graphics data, and appropriate control signals to indicate the end of the frame are also included.

### Color Encoding

One other aspect of the difference between TV standards—and a consequence of the incremental development of television technology (first we had black and white, then color was added—but our black and white TVs could still display a color TV signal—in black and white)—is that color information is encoded in different places in the signal for NTSC and PAL (and SECAM) systems. So, even though the programmer is fully-responsible for controlling the number of scanlines per frame, and the frequency at which frames are generated, it is the Atari itself which encodes the color information into the TV signal.

This is the fundamental reason why there are NTSC, PAL, and SECAM Atari systems—the encoding of the color information for the TV signal! We get some interesting combinations of Atari and games, for example…

If we plug a NTSC cartridge into a PAL '2600, then we know that the NTSC game is generating frames which are 262 lines deep, at 60Hz. But a PAL TV expects frames 312 lines deep, at 50Hz. So the image is only 262/312 of the correct depth, and also images are arriving 60/50 times faster than expected. If we were viewing on a NTSC TV, then the PAL console would be placing the color information for the TV signal in a completely different place than the TV is expecting—so we would see our game in black and white.

There are several combinations you can play with—but the essence is that if you use a different '2600 variant than TV, you will only get black and white (for example, NTSC '2600 with PAL TV or PAL '2600 with NTSC TV) as the color information is not in at the correct frequency band of the signal. And if you plug in a different cartridge than TV (NTSC cart with PAL TV or vice-versa) then what you see depends on the television's capability to synchronize with the signal being generated—as it is not only the incorrect frequency, but also the incorrect number of scanlines.

## Summary

All of this may sound complicated—but really all we need to do is create a 'kernel' (which is the name for your section of an Atari 2600 program which generates the TV frame) which does the drawing correctly—and once that's working, we don't really need to worry too much about the TV—we can abstract that out and just think about what we want to draw.

Well, I lie, but don't want to scare you off TOO early. Wink

Next time, let's have a look how the processor interacts with hardware, I/O and memory.

# Sessions 3 & 6: The TIA and the 6502

Let's spend this session having a look at how some of the hardware generates a scanline for the TV. Remember in session 2, we had a good look at how a TV works, and in particular how a TV frame is composed of 262 scanlines (NTSC) or 312 scanlines (PAL). It's the programmer's job to control how many scanlines are sent to the TV, but it is the '2600 which builds the actual signal comprising the color and intensity information for any scanline. This color and intensity information is derived from the internal 'state' of the TIA (Television Interface Adaptor) chip inside the '2600. The TIA is responsible for creating the signal for a single scanline for the TV.

## The TIA

The TIA 'draws' the pixels on the screen 'on-the-fly'. Each pixel is one 'clock' of the TIA's processing time, and there are exactly 228 color clocks of TIA time on each scanline. But a scanline consists of not only the time it takes to scan the electron beam across the picture tube, but also the time it takes for the beam to return to the start of the next line (the horizontal blank, or retrace). Of the 228 color clocks, 160 are used to draw the pixels on the screen (giving us our maximum horizontal resolution of 160 pixels per line), and 68 are consumed during the retrace period.

### 6502 Clock

The 6502 clock is derived from the TIA clock through a divide-by-three. That is, for every single clock of 6502 time, three clocks of TIA time have passed. Therefore, there are \*exactly\* 228/3 = 76 cycles of 6502 time per scanline. The 6502 and TIA perform a complex 'in-step' dance—one cycle of 6502, three cycles of TIA. A side-note: 76 cycles per line x 262 lines per frame x 60 frames per second = the number of 6502 cycles per second for NTSC (roughly equals 1.19MHz).

So, as our 6502 program is executing its instructions, the TIA is also sending data for each scanline. Every cycle of 6502 time we know that the TIA has sent 3 color clocks of information to the TV. If the TIA was in the first 68 color clocks of the scanline, then it was in the horizontal retrace period. If it was in color clock 68-227, then it was drawing pixels on the visible scanline. And so we go, the 6502 program doing its stuff and at the very same time the TIA doing its stuff.

The magic happens when you start changing the 'state' of the TIA, because those changes are reflected immediately in the TIA output to the TV! Since the 6502 is 'locked' to the TIA through their shared timing origin, it is possible for the programmer to know exactly what pixel on a scanline the TIA is currently drawing. And knowing where the TIA 'is at' allows us to change what it is drawing at particular positions on the scanline. We don't have much scope for change, but we do have some. And it is this ability that master '2600 programmers use to achieve all those amazing effects.

Naturally, to achieve this sort of precision timing, programmers have to know exactly how long the 6502 takes to do each instruction. For example, a load/store combination takes a minimum of 5 cycles of 6502 time. How many onscreen pixels is that? Remember, 3 color clocks per 6502 cycle, so that's 3 x 5 = 15 pixels. Essentially, if one were using the quickest possible load/store combinations to change the color of, say, the background, then the absolute quickest this could be done would be every 15 pixels (just on 11 times per scanline).

### TV Timing

Here's an updated image of the TV timing, taken from the Stella Programming Guide. Some of the numbers should make sense, now. The ones that don't … we'll cover those soon.

Have a good look at this image, and try and understand what it's showing. Your understanding of this will greatly assist your '2600 programming efforts, especially when it comes to designing your kernel.

A screen shot of a video game

Description automatically generated

## Summary

Don't despair! It is not necessary for you to learn how to count 6502 cycles at this stage. Those sort of tricks are for more advanced '2600 programming—and the original design of the TIA hardware made this unnecessary. It's only when you need to push the hardware (TIA) beyond its original design, that you will come to appreciate the benefit inherent in the way that the 6502 and TIA are intricately tied together.

Next session we'll have a closer look at the TIA and how it determines what color to use for each pixel of the scanline it is drawing. In particular, we'll start to look at background, playfield, sprite, missile and ball graphics.

# Session 4: The TIA

Last session we were introduced to the link between the 6502 and the TIA. Specifically, how every cycle of 6502 time corresponds to three color clocks of TIA time.

## The TIA

The TIA determines the color of each pixel based on its current 'state', which contains information about the color, position, size and shape of objects such as background, playfield, sprites (2), missiles (2) and ball. As soon as the TIA completes a scanline (228 cycles, consisting of 160 color clocks of pixels, and 68 color clocks of horizontal blank), it begins drawing the next scanline. Unless there is some change to the TIA's internal 'state' during a scanline, then each scanline will be absolutely identical.

Consequently, the absolute simplest way to 'draw' 262 lines for a NTSC frame is to just WAIT for 262 (lines) x 76 (cycles per line) 6502 cycles. After that time, the TIA will have sent 262 identical lines to the TV. There are other things that we'd need to do to add appropriate control signals to the frame, so that the TV would correctly sync to the frame—but the essential point here is that we can leave the TIA alone and let it do its stuff. Without our intervention, once the TIA is started it will keep sending scanlines (all the same!) to the TV. And all we have to do to draw n scanlines is wait n x 76 cycles.

It's time to have a little introduction to the 6502.

### Binary Numbers

The CPU of the '2600, the 6502, is an 8-bit processor. Basically this means that it is designed to work with numbers 8-binary-bits at a time. An 8-bit binary number has 8 0's or 1's in it, and can represent a decimal number from 0 to 255. Here's a quick low-down on binary…

In our decimal system, each digit 'position' has an intrinsic value. The units position (far right) has a value of 1, the tens position has a value of 10, the hundreds position has a value of one hundred, the thousands position has a value of 1000, etc. This seems silly and obvious—but it's also the same as saying the units position has a value of 10^0 (where ^ means to the power of), the tens position has a value of 10^1, the hundreds position has a value of 10^2, etc. In fact, it's clear to see that position number 'n' (counting right to left, from n=0 as the right-most digit) has a value of 10^n.

That's true of ANY number system, where the 10 is replaced by the 'base'. For example, hexadecimal is just like decimal, except instead of counting 10 digits (0 to 9) we count 16 digits (0 to 15, commonly written 0 1 2 3 4 5 6 7 8 9 A B C D E F—thus 'F' is actually a hex digit with decimal value 15—which again, is 1 x 10^1 + 5 x 10^0 ). So in hexadecimal (or hex, for short), the digit positions are 16^n. There's no difference between hex, decimal, binary, etc., in terms of the interpretation of a number in that number system. Consider the binary number 01100101—this is (reading right to left) … 1 x 2^0 + 0 x 2^1 + 1 x 2^2 + 0 x 2^3 + 0 x 2^4 + 1x2^5 + 1x2^6 + 1x2^7. In decimal, the value is 101. So, %01100101 = 101 where the % represents a binary number. Hexadecimal numbers are prefixed with a $.We'll get used to using binary, decimal and hex interchangeably—after all they are just different ways of writing the same thing. When I'm talking about numbers in various bases, I'll include the appropriate prefix when not base-10.

So now it should be easy to understand WHY an 8-bit binary number can represent decimal values from 0 to 255—the largest binary number with 8 bits would be %11111111—which is 1 x 2^7 + 1 x 2^6 + … + 1 x 2^0.

The 6502 is able to shift 8-bit numbers to and from various locations in memory (referred to as addresses)—each memory location is \*UNIQUELY\* identified by a memory address, which is just like your house street address, or your post-box number. The processor is able to access memory locations and retrieve 8-bit values from, or store 8-bit values to those locations.

### Registers

The processor itself has just three 'registers'. These are internal memory/storage locations. These three registers (named 'A', 'X', and 'Y') are used for manipulating the 8-bit values retrieved from memory locations and for performing whatever calculations are necessary to make your program do its thing.

What can you do with just three registers? Not much … but a hell of a lot of not much adds up to something! Just like with the TV frame generation, a lot of work is left for the programmer. The 6502 cannot multiply or divide. It can only increment, decrement, add and subtract, and it can only work with 8-bit numbers! It can load data from one memory location, do one of those operations on it (if required) and store the data back to memory (possibly in another location). And out of that capability comes all the games we've ever seen on the '2600. Amazing, innit?

At this stage it is probably a good idea for you to start looking for some books on 6502 programming—because that's the ONLY option when programming '2600. Due to the severe time, RAM and ROM constraints, every cycle is precious, every bit is sacred. Only the human mind is currently capable of writing programs as efficiently as required for '2600 development.

That was a bit of a diversion—let's get back to the TIA and how the TIA and 6502 can be used together to draw exactly 262 lines on the TV. Our first task is simply to 'wait' for 76 cycles, times 262 lines.

### NOP

The simplest way to just 'wait' on the 6502 is just to execute a 'nop' instruction. 'nop' stands for no-operation, and it takes exactly two cycles to execute. So if we had 38 'nop's one after the other, the 6502 would finish executing the last one exactly 76 cycles after it started the first. And assuming the first 'nop' started at the beginning of the scanline, then the TIA (which is doing its magic at the same time) would have just finished the last color clock of the scanline at the same time as the last nop finished. In other words, the very next scanline would then start as our 6502 was about to execute the instruction after the last nop, and the TIA was just about to start the horizontal retrace period (which, as we have learned, is 68 color clocks long).

How do we tell the 6502 to execute a 'nop'? Simply typing nop on a line by itself (with at least one leading space) in the source code is all we have to do. The assembler will convert this mnemonic into the actual binary value of the nop instruction. For example…

; sample code

    NOP

    nop

; end of sample code

The above code shows two nop instructions—the assembler is case-insensitive. Comments are preceded by semicolons, and occupy the rest of a line after the ; Opcodes (instructions) are mnemonics—typically 3 letters—and must not start at the beginning of a line! We can have only one opcode on each line. An assembler would convert the above code into a binary file containing two bytes—both $EA (remember, a $ prefix indicates a hexadecimal number) = 234 decimal. When the 6502 retrieves an opcode of $EA, it simply pauses for 2 cycles, and then executes the next instruction. The code sequence above would pause the processor for 4 cycles (which is 12 pixels of TIA time, right?!)

But there are better ways to wait 76 cycles! After all, 38 'nop's would cost us 38 bytes of precious ROM—and if we had to do that 262 times (without looping), that would be 9432 bytes—more than double the space we have for our ENTIRE game!

### WSYNC

The TIA is so closely tied to the 6502 that it has the ability to stop and start the 6502 at will. Funnily enough, at the 6502's will! More correctly, the 6502 has the ability to tell the TIA to stop it (the 6502), and since the TIA automatically re-starts the 6502 at the beginning of every scanline, the very next thing the 6502 knows after telling the TIA to stop the CPU is that the TIA is at the beginning of the very next scanline. In fact, this is the way to synchronize the TIA and 6502 if you're unsure where you're at—simply halt the CPU through the TIA, and next thing you know you're synchronized. It's like a time-warp, or a frozen sleep—you're simply not aware of time passing—you say 'halt' and then continue on as if no halt has happened. It has, but the 6502 doesn't know it.

This CPU-halt is achieved by writing any value to a TIA 'register' called WSYNC. Before we get into reading and writing values to and from 'registers' and 'memory', and what that all means, we'll need to have a look at the memory architecture of the '2600—and how the 6502 interacts with memory, including RAM and ROM.

## Summary

We'll start to explore the memory map (architecture) and the 6502's interaction with memory and hardware, in our next installment.

# Session 5: Memory Architecture

Let's have a look at the memory architecture of the '2600, and how the 6502 communicates with the TIA and other parts of the '2600 hardware.

## Memory Mapping

The 6502 communicates with the TIA by writing, and sometimes reading values to/from TIA 'registers'. These registers are 'mapped' to certain fixed addresses in the 6502's addressing range.

In its simplest form, the 6502 is able to address 65536 (2^16) bytes of memory, each with a unique address. Each 16-bit address ultimately directly controls the 'wires' on a 16-bit buspathway to memory, selecting the appropriate byte of memory to read/write. However, the '2600 CPU, the 6507, is only able to directly access 2^13 bytes (8192 bytes) of memory. That is, only 13 of the 16 address lines are actually connected to physical memory.

This is our first introduction to 'memory mapping' and mirroring. Given that the 6507 can only access addresses using the low 13 bits of an address, what happens if bit 14, 15, or 16 of an address are set? Where does the 6507 go to look for its data? In fact, bits 14,15, and 16 are totally ignored—only the low 13 bits are used to identify the address of the byte to read/write. Consider the valid addresses which can be formed with just 13 bits of data…

from %0000000000000 to %1111111111111

= from $0000 to $1FFF

### Zero is Zero

Note: $0000 is the same as 0 is the same as %000 is the same as %0000000000. 0 is 0. In the same vein, any number with leading zeros is the same as that number without zeros. I often see people writing $02 when they could just write $2, or better yet … 2. Your assembler doesn't care how numbers are written. It's the value of numbers that matter. So use the most readable form of numbers, where it makes sense. Remember, 0 is 0000 is %0 is $000

### Memory Footprint

So we've just written down the minimum and maximum addresses that can be formed with 13 bits. This gives us our memory 'footprint'—the absolute extremes of memory which can be accessed by the 6507 through a 13-bit address.

### Reads and/or Writes

This next idea is important, so make sure you understand! All communication between the CPU and hardware (be it ROM, RAM, I/O, the TIA, or other) is through reads and/or writes to memory locations. Read that again.

The consequences of this are that some of that memory range (between $0 and $1FFF) must contain our RAM, some must contain our ROM (program), and some must presumably allow us to communicate with the TIA and whatever other communication/control systems the machine has. And that's exactly how it works.

### RAM

We have just 128 bytes of RAM on the '2600. That RAM 'lives' at addresses $80 - $FF. It's always there, so any write to location $80 (128 decimal) will actually be to the first byte of RAM. Likewise, any read from those locations is actually reading from RAM.

So we've just learned that the 6507 addresses memory using 13 bits to uniquely identify the memory location, and that some areas of that memory 'range' are devoted to different uses. The area from $80 to $FF is our 128 bytes of RAM!

Don't worry too much about understanding this yet, but TIA registers are mapped in the memory addresses 0 to $7F, RIOT (a bit of '2600 hardware we'll look at later) from $280 - $2FF (roughly), and our program is mapped into address range $1000 to $1FFF (a 4K size).

Note: 1K = 1024 bytes = $400 bytes = %10000000000 bytes.

### The TIA

In essence, then, to change the state of the TIA we just have to write values to TIA 'registers' which look to the 6507 just like any other memory location and which 'live' in addresses 0 to $7F. To the 6502 (and I'll revert to that name now we've emphasized that the 6507 only has 13 address lines as opposed to the 6502's 16 and all other things are equal) a read or write of a TIA register is just the same as a read or write to any other area of memory. The difference is, the TIA is 'watching' those locations, and when you write to that memory, you're really changing the TIA 'registers'—and potentially changing what it draws on a scanline.

## Summary

So now we know how to communicate with the TIA, and where it 'lives' in our memory footprint. And we know how to communicate with RAM, and where it 'lives'. Even our program in ROM is really just another area in our memory 'map'—the program that runs from a cartridge is accessed by the 6502 just by reading memory locations. In effect, the cartridge 'plugs-in' to the 6502 memory map. Let's have a quick look at what we know so far about memory…

|  |  |
| --- | --- |
| Address Range | Function |
| $0000 - $007F | TIA registers |
| $0080 - $00FF | RAM |
| $0200 - $02FF | RIOT registers |
| $1000 - $1FFF | ROM |

We'll keep it simple for now—though you may be wondering what 'lives' in the gaps in that map, between the bits we know about. The short answer is 'not much'—so let's not worry about those areas for now. Just remember that when we're accessing TIA registers, we're really accessing memory from 0 to $7F, and when we access RAM, we're accessing memory from $80 to $FF, etc.

Now that we understand HOW the 6502 communicates with the TIA, one of our next steps will be to start to examine the registers of the TIA and what happens when you modify them. It won't be long now before we start to understand how it all works. Stay tuned.

I might give up writing "next time we'll talk about…" because I seem to end up covering something completely different.

# Session 7: The TV and our Kernel

Time to complete our understanding of what constitutes a TV frame—exactly what has to be sent to the TV to make it display a picture correctly.

Let's take another look at the diagram with the timing information and the Pitfall! image inside.

## TV Timing Diagram

A screen shot of a video game

Description automatically generated

### Numbers Across the Top

Your understanding of the numbers across the top should be good, but we'll briefly revisit what they mean, just to make sure.

There are 228 TIA color clocks on each scanline. 160 of those are spent drawing pixels, and 68 of them are the horizontal retrace period for the TV's scanning of the electron beam back to the start of the next line. In the diagram we see the horizontal blank (retrace) at the left side, so our very first color clock for the TIA's first visible pixel on the screen is cycle #68. We should understand this timing fairly well by now.

### Numbers Down the Right Side

What we're going to finalize this session is our understanding of the numbers down the right-hand side—which represent the scanlines sent to the TV. The diagram shows a valid NTSC TV frame—and thus it consists of 262 scanlines. A PAL diagram would consist of 312 scanlines—and the inner 'picture' area would increase from 192 lines to 242 lines.

#### Reset Signal

Let's go from the top. The first thing that the TV needs is a 'reset signal' to indicate to it that a new frame is starting. This is the 3-scanline section at the very top of the frame. There are special ways to trigger the TIA to send this signal, but we're not going to have to worry too much about understanding that—just about every game does it exactly the same way—all we need to remember is that the first thing to send is that reset trigger (called VSYNC).

#### Vertical Blank and Overscan

TVs are not all made the same. Some cut off more of the picture than others, some show wider pictures, some show taller pictures, etc. To 'standardize' the picture, the diagram shows the recommended spread of valid picture lines, surrounded by blank (or 'overscan') lines. In this case, there are 192 lines of actual picture. We don't \*HAVE\* to stick to this—we could steal some of the lines from the vertical blank section, and some from the overscan section, and increase our picture section appropriately.

As long as our total number of scanlines adds up to 262 for NTSC TVs (or 312 for PAL TVs), then the TV will be able to display the frame. But remember, the further we get 'out of specs' with this method, the less likely it is that ALL TVs will show the picture section in its entirety.

#### Right Side Number Recap

OK, let's march through the numbers on the right side of the diagram.

* 3 Scanlines devoted to the vertical synchronization.
* 37 scanlines of vertical blank time.
* 192 (NTSC) or 242 (PAL) lines of actual picture.
* 30 scanlines of overscan.

Total: 262 scanlines (NTSC) or 312 scanlines (PAL), constituting a valid TV frame. You send the TV this, and it will be a rock-solid display.

### Scanlines and PAL TV

One interesting aside: if you send a PAL TV an \*odd\* number of scanlines, it will only display in black and white. I don't know the exact reason for this, but it must be to do with where/when the color signal is encoded in the TV image, and where the TV looks for it. So remember, always send an even number of scanlines to a PAL TV.

### Scanline Standards

You \*can\* send frames with different numbers of scanlines. That is, 262 and 312 are not totally immutable values. But if you do vary these numbers, it is highly likely that an increasing number of TVs—the further you deviate from these standards—will simply not be able to display your image. So, although you \*can\* … you shouldn't.

Fortunately, emulators available to us today are able to show us the actual number of scanlines which are being generated on each frame. This must have been quite a challenging task for early '2600 programmers—nowadays its quite easy to make sure we get it right.

## Summary

Well, now we have all the knowledge we need about the composition of a TV frame. Once we know how to make the TIA generate its reset signal at the top of the frame, and how to wait the correct amount of time to allow us to correctly generate the right number of scanlines for those other sections, we will be able to design our first 'kernel'—the bit that actually 'draws' the frame.

When we have our kernel working, there's not much more to a '2600 game other than moving sprites around, changing colors, etc. See you next time.

# Session 8: Our First Kernel

We're going to jump right in, now that we know what a kernel needs to do. Seen below, and in the attached file, is the source code for a working '2600 kernel. It displays the image you see here. Not bad for just a few lines of code. Over the next few sessions we'll learn how to modify this code, and assemble it—and, of course, what all those strange words mean.

For now, have a look at the structure of the code and note how closely it relates to the structure of the TV frame diagram in the earlier sessions:

A screen shot of a video game

Description automatically generated

Don't expect to understand everything—we'll walk through every line soon. For now, all you need to know is that the "sta WSYNC" is where the 6502 is telling the TIA to halt the 6502 until the start of the next horizontal blank period (which is at the start of the next scanline, at TIA color clock 0). So each of those lines is where one complete scanline has been sent to the TV by the TIA. Have a close look at those lines, and see how there are 3, followed by 37 (vertical blank period), followed by 192 (picture) followed by 30 (overscan)—and how this exactly matches our TV frame diagram, above.

Yes, this is a complete kernel. It's not that difficult!

## Source Code

            processor 6502

            include "vcs.h"

            include "macro.h"

            SEG

            ORG $F000

Reset

StartOfFrame

   ; Start of vertical blank processing

            lda #0

            sta VBLANK

            lda #2

            sta VSYNC

               ; 3 scanlines of VSYNCH signal...

                sta WSYNC

                sta WSYNC

                sta WSYNC

            lda #0

            sta VSYNC

               ; 37 scanlines of vertical blank...

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

               ; 192 scanlines of picture...

                ldx #0

                REPEAT 192; scanlines

                    inx

                    stx COLUBK

                    sta WSYNC

                REPEND

            lda #%01000010

            sta VBLANK       ; end of screen - enter blanking

               ; 30 scanlines of overscan...

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

                sta WSYNC

            jmp StartOfFrame

            ORG $FFFA

            .word Reset          ; NMI

            .word Reset          ; RESET

            .word Reset          ; IRQ

            END

I tried to make the code sample above as understandable as possible. It is certainly not the most efficient code—for it uses too many bytes of ROM to achieve its effect. But we're learning, and what's important right now is understanding how things work.

Here's a screenshot:

A screenshot of a computer

Description automatically generated

## Summary

Next session we'll have a look at how to actually assemble this code using DASM, and how to make modifications so you can play with it and test it on the emulator to see what effect your changes have.

# Session 9: 6502 and DASM - Assembling the Basics

This session we're going to have a look at the assembler "DASM", what it does, how it does it, why it does it, and how to get it to do it.

The job of an assembler is to convert our source code into a binary image which can be run by the 6502. This conversion process ultimately replaces the mnemonics (the words representing the 6502 instructions we use when writing in assembler) and the symbols (the various names we use for things, such as labels to which we can branch, and various other things like the names of TIA registers, etc) with numerical values.

So ultimately, all the assembler needs to do is figure out a numerical value for all the things which become part of the binary—and place that value in the appropriate place in the binary.

## NOP

We've already had a brief introduction to a 6502 instruction—the one called NOP. This is the no-operation instruction which simply takes 2 cycles to execute. Whenever we enter NOP into our source code, the assembler recognizes this as a 6502 instruction and inserts into the binary the value $EA. This shows that there can be a simple 1:1 relationship between source-code and the binary.

NOP is a single-byte instruction—all it requires is the opcode, and the 6502 will happily execute it. Some instructions require additional 'parameters’ value that is passed to a routine.

## DASM

DASM is the assembler used by most (if not all) modern-day '2600 programmers. It is a multi-platform assembler written in 1988 by Matt Dillon (you should all find his email address and send him a "thank-you" sometime). It's a great tool.

DASM isn't just capable of assembling 6502 (and variant) code—it also has inbuilt capability to assemble code for several other microprocessors. Consequently, one of the very first things that it is necessary to do in our source code is tell DASM what processor the source code is written for…

     processor 6502

This should be just about the first line in any '2600 program you write. If you don't include it, DASM will probably get confused and spit out errors. That's simply because it is trying to assemble your code as if it were written for another processor.

We've just seen how mnemonics (the standard names for instructions) are converted into numerical values by the assembler. Another job the assembler does is convert labels and symbols into values. We've already encountered both of these in our previous sessions, but you may not be familiar with their names.

### Symbol Table

Whenever DASM is doing its job assembling, it keeps a list of all the 'words' it encounters in a file in an internal structure called a symbol table. Think of a symbol as a name for something. Remember the 'sta WSYNC' instruction we used to halt the 6502 and wait for the scanline to be rendered? The 'sta' is the instruction, and 'WSYNC' is a symbol. When it first encounters this symbol, DASM doesn't know much about it, other than what it's called (ie: 'WSYNC'). What DASM needs to do is work out what the \*value\* of that symbol is, so that it can insert that value into the binary file.

When it's assembling, DASM puts all the symbols it finds into its symbol table—and associated with each of these is a value. If it doesn't 'know' the value, that's OK—DASM will keep assembling the rest of the file quite happily. At some point, something in the code might tell DASM what the value for a symbol actually IS—in which case DASM will put that value in its symbol table alongside the symbol. So whenever that symbol is used anywhere, DASM now knows its correct value to put into the binary file.

In fact, it is absolutely necessary for all symbols which go into the binary file to be given values at some point. DASM can't guess values—it's up to you, the programmer, to make sure this happens. A symbol doesn't have to be given a value at any PARTICULAR point in the code, but it does have to be given a value somewhere in the code. DASM will make multiple 'passes'—basically going through the code from beginning to end again and again until it manages to resolve all the symbols to correct values.

#### vcs.h

We've already seen in some sample code how 'sta WSYNC' appears in our binary file as the bytes $85 $02. The first byte $85 is the 'sta' instruction (one variant of many—but let's keep it simple for now) and it is followed by a single byte giving the address of the location into which the byte in the 'A' register is to be stored. We can see this address is location 2 in memory. Somehow, DASM has figured out from the code that the symbol WSYNC has a value of 2, and when it creates the binary file it replaces all occurrences of the symbol with the numeric value 2.

How did it get the value 2? Remember, WSYNC is one of the TIA registers. It appears to the 6502 as a memory location, as the TIA registers are 'mapped' into locations 0 - $7F. The file 'vcs.h' defines (in a roundabout way) the values and names (symbols) for all of the TIA registers. By including the file 'vcs.h' as a part of the assembly for any source file, we automatically tell DASM the correct numeric value for all of the TIA register 'names'.

That's why, at the top of most files, just after the processor statement, we see…

     include "vcs.h"

You don't really need to know much about vcs.h at this stage—but be aware that a 'standardized' version of this file is distributed with the DASM assembler as the '2600 support files package. I would advise you to always use the latest and greatest version of this file. Standards help us all.

So now we know basically what DASM does with symbols—it keeps an internal list of symbols—and their values, if known. DASM will keep going through the code and 'resolving' the symbols into numeric values, until it is complete (or it couldn't find ANYTHING to resolve, in which case it gives an error). Once all symbols have been resolved, your code has been completely processed by the assembler, and it creates the binary image/file for you—and assembly is complete.

### DASM Summary

To summarize: DASM converts source-code consisting of instructions (mnemonics) and symbols into a binary form which can be run by the 6502. The assembler converts mnemonics into opcodes (numbers), and symbols into numbers which it calculates the value of during the assembly process.

### Command Line

DASM is a command-line program—that is, it runs under DOS (or whatever platform you happen to choose, provided you have a runnable version for that platform). DASM is provided with full source-code (it's written in C) so as long as you have a C-compiler handy, you can port it to just about any platform under the sun.

It does come with a manual—and it's always a good idea to familiarize yourself with its capabilities. In the interests of getting you up and running quickly, so you can actually assemble the sample kernel posted a session or two ago, here's what you need to type on the command-line…

dasm kernel.asm -lkernel.txt -f3 -v5 -okernel.bin

This is assuming that the file to assemble is named 'kernel.asm' (.asm is a standard prefix for assembler files, but some prefer to use .s—you can use whatever you want, really, but I always use .asm). Anything prefixed with a minus-sign ('-') is a 'switch'—which tells DASM something about what it is required to do. The -l switch we discussed very briefly, and that tells DASM to create a listing file—in this case, it will write a listing to the file 'kernel.txt'. The -o switch tells DASM what file to use for the output binary—in this case, the binary will be written to 'kernel.bin'. That file can be loaded into an emulator, or burned on an EPROM—it is the ROM file, in other words.

The other switches '-f3' and '-v5' control some internals of DASM—and for now just assume you need these whenever you assemble with DASM. Remember, if you're curious you can always read the manual!

## Output

If all goes well, DASM will output something like this…

DASM V2.20.05, Macro Assembler (C)1988-2003

START OF PASS: 1

----------------------------------------------------------------------

SEGMENT NAME INIT PC INIT RPC FINAL PC FINAL RPC

f000 f000

RIOT [u] 0280 0280

TIA\_REGISTERS\_READ [u] 0000 0000

TIA\_REGISTERS\_WRITE [u] 0000 0000

INITIAL CODE SEGMENT 0000 ???? 0000 ????

----------------------------------------------------------------------

1 references to unknown symbols.

0 events requiring another assembler pass.

--- Symbol List (sorted by symbol)

AUDC0 0015

AUDC1 0016

AUDF0 0017

AUDF1 0018

AUDV0 0019

AUDV1 001a

COLUBK 0009 (R )

COLUP0 0006

COLUP1 0007

COLUPF 0008

CTRLPF 000a

CXBLPF 0006

CXCLR 002c

CXM0FB 0004

CXM0P 0000

CXM1FB 0005

CXM1P 0001

CXP0FB 0002

CXP1FB 0003

CXPPMM 0007

ENABL 001f

ENAM0 001d

ENAM1 001e

GRP0 001b

GRP1 001c

HMBL 0024

HMCLR 002b

HMM0 0022

HMM1 0023

HMOVE 002a

HMP0 0020

HMP1 0021

INPT0 0008

INPT1 0009

INPT2 000a

INPT3 000b

INPT4 000c

INPT5 000d

INTIM 0284

NUSIZ0 0004

NUSIZ1 0005

Overscan f02c (R )

PF0 000d

PF1 000e

PF2 000f

Picture f01d (R )

REFP0 000b

REFP1 000c

RESBL 0014

Reset f000 (R )

RESM0 0012

RESM1 0013

RESMP0 0028

RESMP1 0029

RESP0 0010

RESP1 0011

RSYNC 0003

StartOfFrame f000 (R )

SWACNT 0281

SWBCNT 0283

SWCHA 0280

SWCHB 0282

T1024T 0297

TIA\_BASE\_ADDRESS 0000 (R )

TIM1T 0294

TIM64T 0296

TIM8T 0295

TIMINT 0285

VBLANK 0001 (R )

VDELBL 0027

VDELP0 0025

VDELP1 0026

VerticalBlank f014 (R )

VSYNC 0000 (R )

WSYNC 0002 (R )

--- End of Symbol List.

Complete.

Here we can actually SEE the symbol table, and the numeric values that DASM has assigned to the symbols. If you look at the listing file, wherever any of these symbols is used, you will see the corresponding number in the symbol table has been inserted into the binary.

There are lots of symbols there, as the vcs.h file defines just about everything you'll ever need to do with the TIA. The symbols which are actually USED in your code are marked with a (R )—indicating 'referenced'.

Now you should be able to go and assemble the sample kernel I provided earlier. Don't be afraid to have a play with things, and see what happens! Experimenting is a big part of learning.

## Summary

Soon we'll start playing with some TIA registers and seeing what happens to our screen when we do that! For now, though, make sure you are able to assemble and run the first kernel. If you have any problems, ask for assistance and I'm sure somebody will leap to your aid.

# Session 10: Orgasm