Data in Program Space

Introduction

So you have some constant data and you're running out of room to store it? Many AVRs have limited amount of RAM in which to store data, but may have more Flash space available. The AVR is a Harvard architecture processor, where Flash is used for the program, RAM is used for data, and they each have separate address spaces. It is a challenge to get constant data to be stored in the Program Space, and to retrieve that data to use it in the AVR application.

The problem is exacerbated by the fact that the C Language was not designed for Harvard architectures, it was designed for Von Neumann architectures where code and data exist in the same address space. This means that any compiler for a Harvard architecture processor, like the AVR, has to use other means to operate with separate address spaces.

Some compilers use non-standard C language keywords, or they extend the standard syntax in ways that are non-standard. The AVR toolset takes a different approach.

GCC has a special keyword, `__attribute__` that is used to attach different attributes to things such as function declarations, variables, and types. This keyword is followed by an attribute specification in double parentheses. In AVR GCC, there is a special attribute called `progmem`. This attribute is use on data declarations, and tells the compiler to place the data in the Program Memory (Flash).

AVR-Libc provides a simple macro `PROGMEM` that is defined as the attribute syntax of GCC with the `progmem` attribute. This macro was created as a convenience to the end user, as we will see below. The `PROGMEM` macro is defined in the <**avr/pgmspace.h**> system header file.

It is difficult to modify GCC to create new extensions to the C language syntax, so instead, avr-libc has created macros to retrieve the data from the Program Space. These macros are also found in the <**avr/pgmspace.h**> system header file.

A Note On const

Many users bring up the idea of using C's keyword `const` as a means of declaring data to be in Program Space. Doing this would be an abuse of the intended meaning of the `const` keyword.

`const` is used to tell the compiler that the data is to be "read-only". It is used to help make it easier for the compiler to make certain transformations, or to help the compiler check for incorrect usage of those variables.

For example, the const keyword is commonly used in many functions as a modifier on the parameter type. This tells the compiler that the function will only use the parameter as read-only and will not modify the contents of the parameter variable.

`const` was intended for uses such as this, not as a means to identify where the data should be stored. If it were used as a means to define data storage, then it loses its correct meaning (changes its semantics) in other situations such as in the function parameter example.

Storing and Retrieving Data in the Program Space

Let's say you have some global data:

```
unsigned char mydata[11][10] =
{
    {0x00,0x01,0x02,0x03,0x04,0x05,0x06,0x07,0x08,0x09},
    {0x0A,0x0B,0x0C,0x0D,0x0E,0x0F,0x10,0x11,0x12,0x13},
    {0x14,0x15,0x16,0x17,0x18,0x19,0x1A,0x1B,0x1C,0x1D},
    {0x1E,0x1F,0x20,0x21,0x22,0x23,0x24,0x25,0x26,0x27},
    {0x28,0x29,0x2A,0x2B,0x2C,0x2D,0x2E,0x2F,0x30,0x31},
    {0x32,0x33,0x34,0x35,0x36,0x37,0x38,0x39,0x3A,0x3B},
    {0x3C,0x3D,0x3E,0x3F,0x40,0x41,0x42,0x43,0x44,0x45},
    {0x46,0x47,0x48,0x49,0x4A,0x4B,0x4C,0x4D,0x4E,0x4F},
    {0x50,0x51,0x52,0x53,0x54,0x55,0x56,0x57,0x58,0x59},
    {0x5A,0x5B,0x5C,0x5D,0x5E,0x5F,0x60,0x61,0x62,0x63},
    {0x64,0x65,0x66,0x67,0x68,0x69,0x6A,0x6B,0x6C,0x6D}
};
```

and later in your code you access this data in a function and store a single byte into a variable like so:

```
byte = mydata[i][j];
```

Now you want to store your data in Program Memory. Use the `PROGMEM` macro found in <**avr/pgmspace.h**> and put it after the declaration of the variable, but before the initializer, like so:

```
#include <avr/pgmspace.h>
.
.
.
unsigned char mydata[11][10] PROGMEM =
{
    {0x00,0x01,0x02,0x03,0x04,0x05,0x06,0x07,0x08,0x09},
    {0x0A,0x0B,0x0C,0x0D,0x0E,0x0F,0x10,0x11,0x12,0x13},
    {0x14,0x15,0x16,0x17,0x18,0x19,0x1A,0x1B,0x1C,0x1D},
    {0x1E,0x1F,0x20,0x21,0x22,0x23,0x24,0x25,0x26,0x27},
    {0x28,0x29,0x2A,0x2B,0x2C,0x2D,0x2E,0x2F,0x30,0x31},
    {0x32,0x33,0x34,0x35,0x36,0x37,0x38,0x39,0x3A,0x3B},
    {0x3C,0x3D,0x3E,0x3F,0x40,0x41,0x42,0x43,0x44,0x45},
    {0x46,0x47,0x48,0x49,0x4A,0x4B,0x4C,0x4D,0x4E,0x4F},
    {0x50,0x51,0x52,0x53,0x54,0x55,0x56,0x57,0x58,0x59},
    {0x5A,0x5B,0x5C,0x5D,0x5E,0x5F,0x60,0x61,0x62,0x63},
    {0x64,0x65,0x66,0x67,0x68,0x69,0x6A,0x6B,0x6C,0x6D}
};
```

That's it! Now your data is in the Program Space. You can compile, link, and check the map file to verify that `mydata` is placed in the correct section.

Now that your data resides in the Program Space, your code to access (read) the data will no longer work. The code that gets generated will retrieve the data that is located at the address of the `mydata` array, plus offsets indexed by the `i` and `j` variables. However, the final address that is calculated where to the retrieve

the data points to the Data Space! Not the Program Space where the data is actually located. It is likely that you will be retrieving some garbage. The problem is that AVR GCC does not intrinsically know that the data resides in the Program Space.

The solution is fairly simple. The "rule of thumb" for accessing data stored in the Program Space is to access the data as you normally would (as if the variable is stored in Data Space), like so:

```
byte = mydata[i][j];
```

then take the address of the data:

```
byte = &(mydata[i][j]);
```

then use the appropriate `pgm_read_*` macro, and the address of your data becomes the parameter to that macro:

```
byte = pgm_read_byte(&(mydata[i][j]));
```

The `pgm_read_*` macros take an address that points to the Program Space, and retrieves the data that is stored at that address. This is why you take the address of the offset into the array. This address becomes the parameter to the macro so it can generate the correct code to retrieve the data from the Program Space. There are different `pgm_read_*` macros to read different sizes of data at the address given.

Storing and Retrieving Strings in the Program Space

Now that you can successfully store and retrieve simple data from Program Space you want to store and retrive strings from Program Space. And specifically you want to store and array of strings to Program Space. So you start off with your array, like so:

```
char *string_table[] =
{
    "String 1",
    "String 2",
    "String 3",
    "String 4",
    "String 5"
};
```

and then you add your PROGMEM macro to the end of the declaration:

```
char *string_table[] PROGMEM =
{
    "String 1",
    "String 2",
    "String 3",
    "String 4",
    "String 5"
};
```

Right? WRONG!

Unfortunately, with GCC attributes, they affect only the declaration that they are attached to. So in this case, we successfully put the `string_table` variable, the array itself, in the Program Space. This DOES NOT put the actual strings themselves into Program Space. At this point, the strings are still in the Data Space, which is probably not what you want.

In order to put the strings in Program Space, you have to have explicit declarations for each string, and put each string in Program Space:

```
char string_1[] PROGMEM = "String 1";
char string_2[] PROGMEM = "String 2";
char string_3[] PROGMEM = "String 3";
char string_4[] PROGMEM = "String 4";
char string_5[] PROGMEM = "String 5";
```

Then use the new symbols in your table, like so:

```
PGM_P string_table[] PROGMEM =
{
    string_1,
    string_2,
    string_3,
    string_4,
    string_5
};
```

Now this has the effect of putting `string_table` in Program Space, where `string_table` is an array of pointers to characters (strings), where each pointer is a pointer to the Program Space, where each string is also stored.

The `PGM_P` type above is also a macro that defined as a pointer to a character in the Program Space.

Retrieving the strings are a different matter. You probably don't want to pull the string out of Program Space, byte by byte, using the **`pgm_read_byte()`** macro. There are other functions declared in the <**avr/pgmspace.h**> header file that work with strings that are stored in the Program Space.

For example if you want to copy the string from Program Space to a buffer in RAM (like an automatic variable inside a function, that is allocated on the stack), you can do this:

```
void foo(void)
{
    char buffer[10];

    for (unsigned char i = 0; i < 5; i++)
    {
        strcpy_P(buffer, (PGM_P)pgm_read_word(&(string_table[i])));

        // Display buffer on LCD.
    }
    return;
}
```

Here, the `string_table` array is stored in Program Space, so we access it normally, as if were stored in Data Space, then take the address of the location we want to access, and use the address as a parameter to `pgm_read_word`. We use the `pgm_read_word` macro to read the string pointer out of the `string_table`

array. Remember that a pointer is 16-bits, or word size. The `pgm_read_word` macro will return a 16-bit unsigned integer. We then have to typecast it as a true pointer to program memory, `PGM_P`. This pointer is an address in Program Space pointing to the string that we want to copy. This pointer is then used as a parameter to the function `strcpy_P`. The function `strcpy_P` is just like the regular `strcpy` function, except that it copies a string from Program Space (the second parameter) to a buffer in the Data Space (the first parameter).

There are many string functions available that work with strings located in Program Space. All of these special string functions have a suffix of `_P` in the function name, and are declared in the <**avr/pgmspace.h**> header file.

Caveats

The macros and functions used to retrieve data from the Program Space have to generate some extra code in order to actually load the data from the Program Space. This incurs some extra overhead in terms of code space (extra opcodes) and execution time. Usually, both the space and time overhead is minimal compared to the space savings of putting data in Program Space. But you should be aware of this so you can minimize the number of calls within a single function that gets the same piece of data from Program Space. It is always instructive to look at the resulting disassembly from the compiler.
