RFSH pin and the refresh patent

The RFSH pin is driven by a simple circuit. A gate is active for M1T3 or M1T4. This goes through a latch that is active when the clock is high. This latch delays the signal by half a clock (since M and T lines change on the low clock). Thus, the RFSH pin is low during M1 for T3h, T4l, T4h, and T1l.

Note that the RFSH pin is not inactivated by BUSREQ/BUSACK. However, refreshes will not take place while the Z80 is disabled by BUSREQ. Thus, a long BUSREQ time will block refresh, which may cause data loss.

Building a memory refresh feature into the CPU is one of the innovative features of the Z80. The refresh feature of the Z80 is patented in https://www.google.com/patents/US4332008. This patent was filed by Shima, Faggin, and Ungermann in 1979.

The patent specifically covers putting the PC value on the address bus in the first period of a fetch cycle, and then in the second period of the fetch cycle acting on the data as well as putting the the refresh counter on the address bus in the fetch cycle, and incrementing both the PC and refresh counter before the next cycle.

The second claim is similar, covering fetch in one time period, and then refreshing in a second time period while the CPU is processing the instruction and doesn't use the address bus.

The patent goes into a lot of detail on a typical Z80 system, the T state timing, etc.

Figure 14 shows the timing of the address pins: the PC is loaded into the address latch in T1, the incremented PC is stored back in T2, the refresh register goes to the address latch in T3, and the incremented refresh register is stored back in T4. Meanwhile, the accumulator and flags are loaded in T3, the second ALU argument is loaded from registers in T4, and the ALU results are stored back to the accumulator and flags in T1.

Figure 15 in particular is relevant. It shows the Z80's address and data bus architecture in detail. Of interest is the partitioning of the data bus into multiple parts: the data pins and Instruction Register, the lower half registers including flags and the associated flags in the ALU, the upper half registers, and the ALU bus. In addition, the PC and refresh register are partitioned from the rest of the registers. This data bus structure will be discussed in more detail in the data bus chapter.

The patent goes into pages of detail describing the Z80's overall architecture, register set, instructions, and so forth. Unlike many patents which are written in legalese, this patent gives a well-written description of the Z80.

Of relevance to this chapter is the discussion of dynamic RAM. Dynamic RAM requires periodic refreshing of its contents to prevent loss of data. This is required because the data is stored in capacitors that gradually lose their charge, rather than in stable latches. The advantage of this approach is the memory cells are much smaller than static RAM, since the multiple transistors of the latch are avoided. Thus, more dynamic RAM can fit on a chip and the price is lower.

To refresh a dynamic RAM, the rows of the RAM must be addressed. The corresponding memory cells are read out, amplified, and stored back. As the patent explains, previous systems have elaborate external logic to perform the refresh. This typically consists of a refresh counter that steps through the necessary addresses.

The Z80 performs refreshes in a transparent way that doesn't affect CPU speed, since the refreshes are done in timing cycles where the address lines are not otherwise in use. In particular, the Z80 uses the cycles after an instruction fetch while the instruction is being decoded.

The patent describes typical memory of the time, where 8 kilobytes of memory is implemented using 18-pin 4K memory chips, in particular an array of 16 chips, arranged in two banks of 8 chips. Each 4K chip requires 12 address lines, of which 7 are used for refresh. One disadvantage of the Z80's refresh scheme is that only 7 bits of refresh address are available. As chips

got larger, the 7-bits of refresh became insufficient.

The hardware reason that the R register only provides 7 bits of refresh and the top bit isn't incremented comes from the structure of the 16-bit incrementer. Normally, the incrementer increments all 16 bits. But for refresh, obviously the I register should not be incremented. Due to the partitioned structure of the incrementer, when incrementing is restricted, only 7 bits are incremented, instead of the full byte that you would expect. See the chapter on the incrementer for details.

The 7-bit refresh was sufficient for 64Kx1 chips such as the 4164, even though they use 8 address bits. However, a 256Kx1 such as the 41256 required 8 bit refresh. The Z80's 7-bit refresh was also not sufficient for a 64Kx4 DRAM such as the 41464.
http://www.datasheets360.com/pdf/8558050608332472 43
The 41464 has 8 address bits and requires 256 refresh cycles (over 4 ms).
The 41256 was organized as 256Kx1 with 8 multiplexed address lines. It required 256 refresh cycles, which the Z80 couldn't provide with its 7 bits.

It may not be well-known that Zilog produced memory chips in this era. The databook (http://www.classiccmp.org/cini/pdf/Zilog/Zilog%20Data%20Book.PDF) from approximately 1979 describes a series of memory chips that Zilog produced:
The Z6104 was a static 4Kx1 bit chip.
The Z6114 was a static 1Kx4 bit chip with I/O lines. The Z6142 variant had additional enable inputs for additional control.
The Z6116 was a 16x1 bit dynamic RAM. See the datasheet http://www.datasheets360.com/pdf/9079771206256639178.

It's interesting to note that one of the motivations for Faggin to leave Intel was that Intel was more focused on DRAM production than CPUs. But then Zilog started producing DRAMs as well. But around 1983, Japanese manufacturers became able to produce inexpensive DRAMs, causing Intel to focus on microprocessors instead of memory, as memory became less profitable to produce.

Note that the dynamic RAM chip provided 4 times the storage capacity of the static RAM chips. The static RAM chips had access times from 150 ns to 350 ns, with cycle times from 240 ns to 510 ns. The Z6116 has access times from 150 to 250 ns, with read-modify-write cycle times from 320 to 500 ns.

The Z6116 must have all the rows refreshed within 2 ms (128 rows). This ensures that the capacitor will not have lost its charge during this time. (This is the same refresh time as the Intel 2107 requires.)

The chips are described as being manufactured with n-channel depletion-load silicon-gate technology using ion-implantation and shallow junctions for high performance. They were projection printed on four-inch wafers for high yield and low cost. (For comparison, current wafer sizes are 300mm (almost 12 inches) with plans to move to 450mm wafers (17.7 inches). The latter size is almost 20 times the size of a four-inch wafer, showing the huge increase in wafer sizes.

Interestingly, the typist of the patent got confused and typed the HL register as the III register.

The patent discusses previous techniques for refreshing memory such as the Intel 3222 refresh controller, which provides 6 bits of refresh, designed for the Intel 2107B memory.
http://www.andysarcade.de/data/electronics/components/2107a.pdf This is a 4Kx1 dynamic RAM . The 3222 refresh controller contained a 6-bit counter, along with a multiplexer that selected between the 6-bit address lines from the computer and the refresh lines and provided the appropriate lines to the memory.

Avoiding the additional chips required for external refresh, such as with the Intel 3222, is an advantage of the Z80. One of the motivations of the Z80 was to minimize the number of additional chips required, so providing built-in refresh helps achieve this goal.