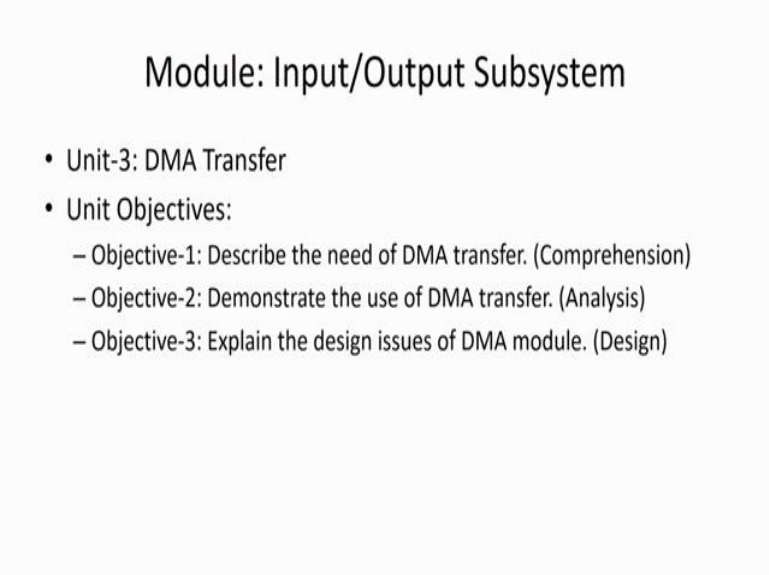
Computer Organization and Architecture: A Pedagogical Aspect

Prof. Jatindra Kr. Deka
Dr. Santosh Biswas
Dr. Arnab Sarkar
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati

Lecture – 35
DMA Transfer

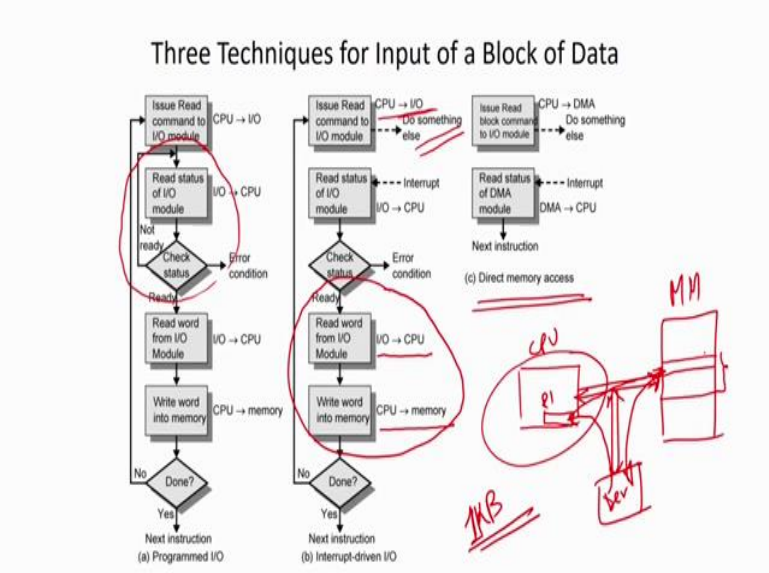
Hello everybody, welcome back to the online course on computer organization and architecture. We are in the input output subsystem. Now unit three is related to DMA transfer, already we have said that there are three ways to transfer information programmed I/O. interrupt driven I/O and DMA transfer. Already we have discussed about the other two issues. Now we are coming to the third mode of transfer this is your DMA transfer. What is the objective of this particular module DMA transfer.

(Refer Slide Time: 01:00)



So, objective 1: describe the need of DMA transfer, it will be in the comprehensive level objective 1: demonstrate the use of DMA transfer. This will be discussed in the analysis level, explain the design issues of DMA module. So, it will be in the design level. So, we are going to see what are issues that are, that need to be discussed when we are going to design a DMA module direct memory access, DMA is direct memory access. So, already you just see that what we did in the case of your programmed I/O.

(Refer Slide Time: 01:36)



We have problem over here when processor is having busy waiting. So, to overcome these things we have come with the interrupt driven I/O. So, we have eliminated this particular busy waiting and after initiating the transfer now processor can do something else, but here also if you look into the complete transfer process what you will find that, in this particular during transfer if you look into it then what will happen? If it is coming from I/O to processor.

Basically processors are having some registers, temporary storage and if we are going to keep more information, transfer more information. Finally, we are going to keep it into the main memory. You just see that when device is ready to transfer information and what will happen? First we are transferring it from devices to the CPU; that means, we are bring it to some processor register and from processor register we are transferring it to memory.

So, basically what will happen? You just see that if this is my processor CPU and this is my main memory ok, main memory and it is connected to system bus and this through system bus; say we are connecting to a device. So, say that I am now eliminating the I/O module say the device is direct connected, then in this particular mode of operation what will happen.

If we want to transfer some information from some device, may be, say if we are, if we are trying to transfer a file from your hard disk to your main memory, then we are storing this particular information in hard disk and we are going to bring it to the main memory, just say that if file size of say 1 kilobyte, then what will happen, and we are going to bring this particular

1 kilobyte to the memory, because we don't have a storage space inside the processor that one kilobyte storage space. We are having hardly pure registers, may be 8 to 16.

So, in this particular way what it is doing from device first it will come to some register inside the processor may be say register R_1. So, it is coming to the register and from register we are storing it to the memory location. So, during this data transfer operation, processor is involved, processor is actively involved during the data transfer. So, when we are going to transfer 1 kilobyte of information then what will happen.

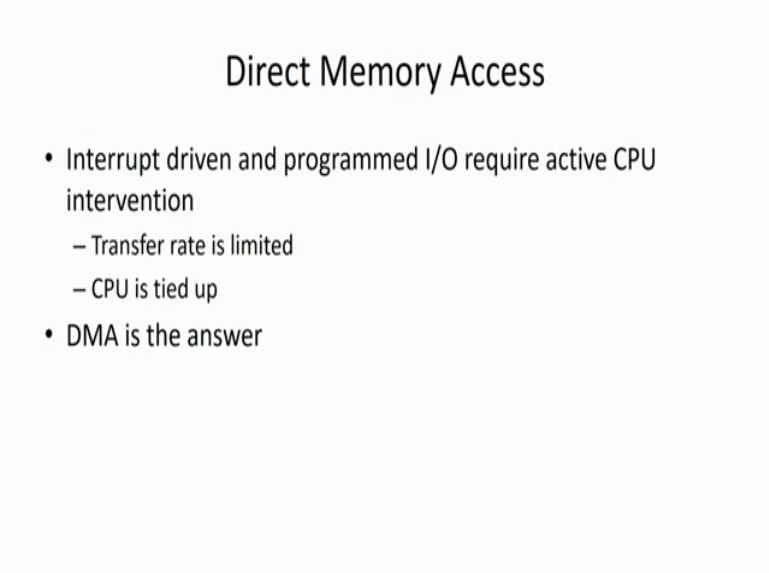
We are going to transfer it byte by byte and no maybe say if it is a 16 bit configuration maybe 2 byte at a time like that we are going to transfer it and for transferring enter 1 kilobyte, what will happen? the processor is involved, processor cannot do any other work, because it is taking the information from device, bringing it to the processor register, from processor register we are transferring into the memory. So, processor is mostly involved in transferring the information.

So, we should think about that you know. So, where the processor can be freed while doing the data transfer. So that processor can carry out some other work. So, for that the solution is your DMA; direct memory access. So, in case of direct memory access what will happen. The data transfer will take place between device and memory, involvement of processor will be eliminated.

So, this is the basic crux about the DMA, and why you are coming to DMA. This is the main reason that we want to remove the role of the processor during the data transfer. So, if we can directly transfer information from device to memory, during that time processor, may carry out some other work, but what are the things processor can carry out, we will see that thing also ok.

So, this is the basic concept, why we are coming for the DMA direct memory access. So, DMA says direct memory access and data transfer takes place between memory and devices the intervention of processor is eliminated.

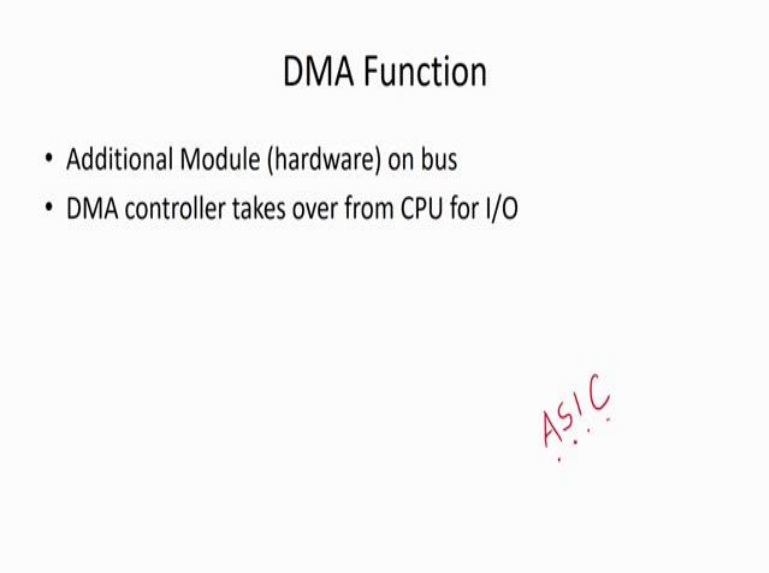
(Refer Slide Time: 05:47)



So, this is basically already explained interrupt driven and program I/O requires the active CPU intervention; that means, CPU is always engaged while transferring the information, transfer rate is limited and CPU is tied up. So, it cannot do any other work, just to eliminate all those things, what is the answer? DMA is the transfer. So, direct memory access we can look for it. So, that the overhead; that is getting by the processor can be eliminated if we are going to use DMA.

Now we are going to see what are the issues while we are going to design the DMA, DMA controller direct memory access. And I think after the completion of this course you will be having an idea how DMA works, even you will be in a position to design a new DMA controller.

(Refer Slide Time: 06:38)



So, for that what will happen? We are having an additional module in the hardware or we are connecting to the bus, it is known as your DMA controller and when we are going to have transfer, then this DMA controller. So, you just see that we are having I/O module that I/O module can transfer information. So, DMA controller can be also treated as an I/O module only, it is an input output module, but it is having a specific functionality.

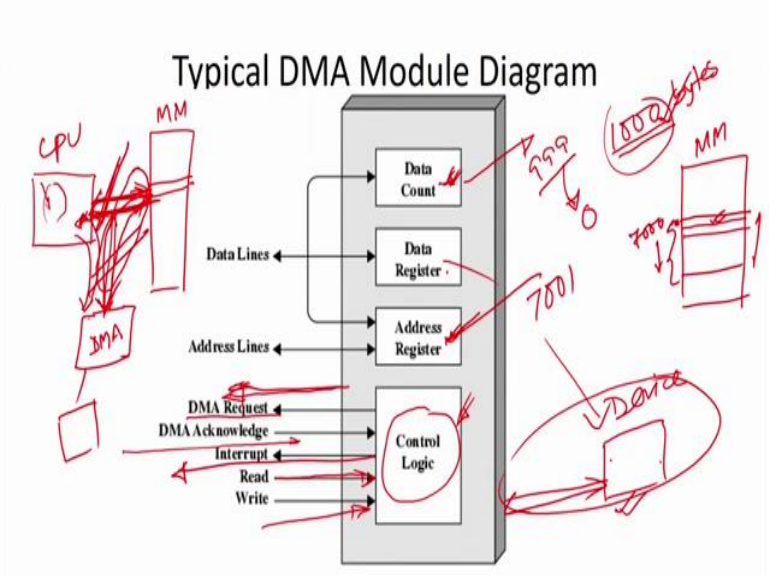
So, in that particular case what will happen. Now DMA controller going to take over the data transfer from CPU and it will carry out this; that means, it is having some processing tasks even; that means, you can think about that we are designing another processor dedicated processor which can carry out a specific tasks only. What specific task is can do. It can do the data transfer between device and processor.

So, I think I at some point up time, I have mentioned about ASIC application specific integrated circuit. So, DMA may fall in this particular category. In this category of IC's ASIC application specific integrated circuit and what is that application the basic application that DMA is going to handle, is your data transfer between I/O devices and memory and how why you require these things, because first of all you have to bring the information to the memory and how processor works? Processor works on Von Neumann stored program principle and processor is going to access the information from main memory which is the storage in this particular case.

So, see processor works on Von Neumann stored program principle. So, when processor is going to execute some program or going to carry out some tasks, it is going to get the information from main memory which is basically storage unit. Now how we are going to get the information to the storage unit from input output devices. So, for that, now to transfer information from input devices to the main memory we can say that we are having a device controller called DMA controller which is an ASIC application specific IC and the tasks perform by DMA controller is transferring information from input devices to the memory. On the other hand along with that you can say that transferring information from memory to the output devices, because the result we have to give it to the users.

So, DMA controller is coming in between and it is going to transfer the information from devices to the memory all right. So, this is a function of a DMA controller and we can view it as ASIC application specific integrated circuit.

(Refer Slide Time: 09:26)



Now, the typical DMA module diagram, you just see what are the things that we are having. So, these are that issues that whatever is relevant just I am keeping it here, but along with that this addressing techniques and other things are similar to our I/O module, because we have to have a status register, we have to give the addresses, so that we can identify the devices. All those things are there in DMA controller also, but along with that we are having some specific components.

So, what are the specific component one is your control logic ok. So, this is the control logic after getting the signals from the output environment; that means, what is the output environment in that particular case. Now output environment is processor, it is going to get some indication, some signal from the processor along with that, it is going to get some information or signals from the devices connected to the DMA controller ok.

So, in the DMA model, those are the signals that it is going to get and depending on those particular signal this control logic, we need to design this particular control logic to carry out the appropriate task and to carry out the complete operation of transferring the information from devices to the memory or from memory to the devices. So, this is the control logic we need to design. So, when we are going to design it, it will be a simple control logic after knowing the all the input signals, all the output signal and the way we know what as the things that it is going to perform, depending on that we can design this particular control logic.

So, it is a simple control logic. I think by this time you know how you are going to design a control unit. So, control unit is a part of my processor CPU, because CPU is having some registers ALU and control unit and in this particular subject, in this particular course, we have seen the design issues of this particular control unit to synchronize operation of the processor. So, with respect to that control unit, this control logic is a very simple one. If you put your time, if you put slight mind what need to be done, you can design this particular control logic also.

So, what are the inputs and output to this control logic; one signal is your read or write, you are saying two signals. So, if it is a read; that means, you are going to take information from devices to the memory then it is an read information. If we are going to put information from memory to the output devices, it is write. So, this signal will come from the processor. It will give indicate the process will say whether it is a read or it is a write.

Along with that we are going to get other information also from which devices we are going to get it, we are not showing it over here. So, that information, also it will be given and accordingly this DMA module is going to work with that particular device just say that I am just writing it as a some devices there. So, this device is also connected to the I/O module. This portion we have not shown in the wire. So, basically it is giving the addresses and the status line of these particular devices.

So, once, it DMA module is getting the information that it is a read information and it is going to read from a particular device that after getting this information it will pass these things to the device and going to collect the appropriate thing and going to set it up. So, once it will set it up that this device is ready for the data transfer and other things then what will happen? That DMA controller or DMA module will give DMA request to the processor. So, processor is connected over here.

Now once it is giving that DMA request then it will give DMA acknowledgement, but when it will give DMA acknowledgment, when the processor is ready to perform the DMA transfer, now what it is going to do. So, one it is coming that DMA device is ready and everything is said, it will give the DMA request in between. Now processor, what processor is going to do. It is at least we must know what is the volume of the data we are going to transfer.

So, it will be in the, you are having a data count registers, just say that for a simple example I am saying that we are going to transfer say 1000 byte of information ok. So, this count will be set to 1000 ok. So, say that we are going to transfer 1000 byte of information. This is the requirement along with that now processor is going to say that it is going to read it.

Now after get bringing this information where we are going to store it. So, I am having a memory, main memory; say I am going to store this particular information from starting at this say 7000 onward. So, what will happen? that processor is going to make it ready that address register will be set to this particular 7000. Now just see what does it means?

We are going to transfer 1000 bytes of information; say it is byte organized and I am in memory also in, we are going to store 1 byte of information and where we are going to store it up, bringing it from the devices say from memory location 7000. So, this addresses will be set to the 7000.

So, this is the address line which is connected to the processor. This is the data line connected to the processor. So, through this data bus we are setting this data count to the 1000. Through this data line we are setting it to the address start to 7000. Now everything is set. Now device is ready, it will be going to give this DMA request once processor is ready then processor will give the DMA acknowledgement.

Now what will happen in that particular case when the DMA acknowledgement is coming over here at the particular point, that DMA controller is going to take the control of the bus.

Now you just see what will happen? Say this is my processor CPU, this is the memory ok. This is connected through the system bus and say this I am showing DMA controller is connected to it ok. Now say processor is initiating it, it is saying that after setting this data count and your starting address.

Now processor is initiating that it wants to perform a read operation ok. So, in that particular time that control logic, we have designed in appropriate way. Now it is going to look for the appropriate devices, because addresses will also come from these particular devices, from which devices we are going to read it. So, once everything is set. Now device is ready, we are at the point of transferring the information then processor is giving the DMA request ok. It is having that getting the DMA request. Now when DMA request is coming, now processor is going to say that. Now we can perform the transfer operation give the DMA acknowledgement.

Then when DMA acknowledgment is coming for the DMA at the particular time, this control of the bus will be given to the DMA; that means, now processor is slightly dealing through control signal we are setting it and at the particular point now we are having this particular connection. Now processor is slightly dealing. Now processor is not going to use the system bus, system bus will be used by the DMA controller. Now what it will do? Now it will take the information from device to the data register.

So, this is the DMA module from data register, it will going to store into the memory location ok. Once 1 byte of information is transferred; that means, you are the getting the information from device, through data register it is going to the memory location then what will happen? Then data count will be decremented; that means, it will become say 999; that means, already have transfer one, we need to transfer 999 bytes and this address register will be incremented and it will say that 7001; that means, first byte we have stored in the register memory location 7000. Next byte we need to store in 7001.

So, like that data count will be decremented after every transfer and address register will be incremented just to point to the next memory location. So, when we transferred that 1000 memory byte; that means, when this value will come to 0; that means, we have transferred all the 1000 byte of information. So, after completion of the transfer, now this particular DMA controller will issue this particular interrupt signal to the processor.

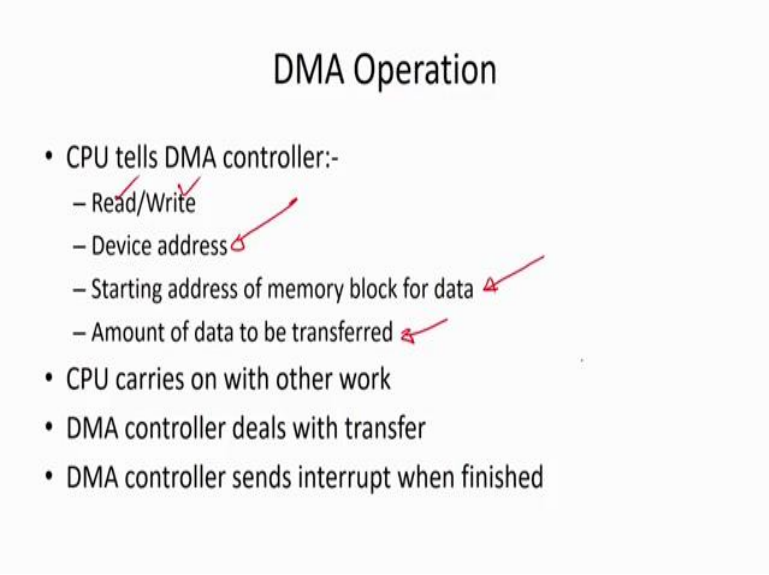
So, this interrupt signal will say that now transfer is over, it is completed. Now what it is going to do? once it get this particular signal then processor is going to take the control of the system

bus; that means, now system bus will be connected to this particular processor and now this is the link. Now DMA is not directly connected to the system bus. So, this is the way that DMA is going to transfer information from devices to the memory.

So, what is the basic principle over here. Basically DMA is going to take control of the system bus and in system bus, basically it is going to look for address bus and data bus and thereby carry out the transfer once it completes the operation, it will give an indication to the processor, then processor is going to take back the system bus; that means, now DMA controller is no longer connected to the DMA bus, that now processor can work with the main memory through this particular system bus.

So, this is the way we are transferring information from devices to the memory through DMA controller. And similarly from memory to the devices also we can transfer by following the same principle, except that here we are having this particular write signal and that will be transfer from memory to the devices through this particular data register. So, these are the basic components that we have in a DMA module or DMA controller and it works with the help of those control signals ok. This is the way that DMA controller works.

(Refer Slide Time: 20:16)



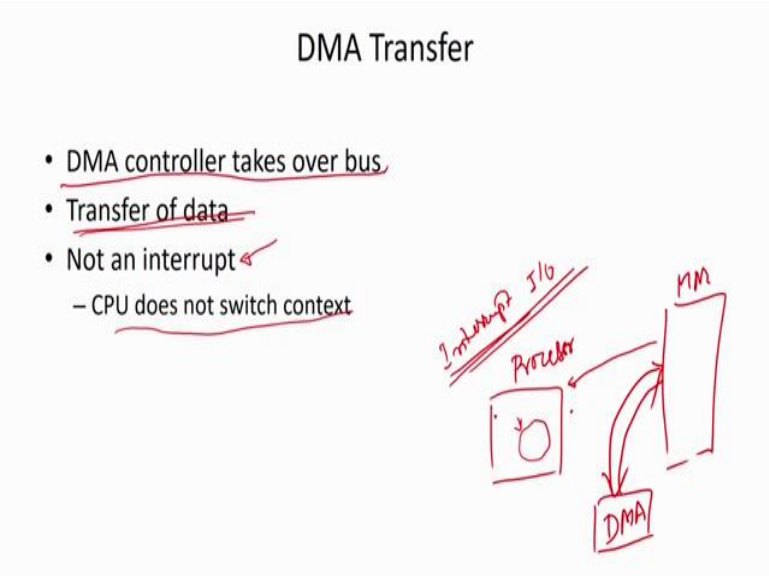
So, basically now we know how DMA controller is going to work, how it transfers the information without intervening the processor operation that means processor is not involved at a, not involved in this particular data transfer. So, basically now what are the DMA operations, just see in that particular case, CPU tells the DMA controller whether it is read or

write. This is the information that processor will give. It will give that device addresses also from which device it is going to take the information, starting address of the memory block of data.

So, it will give the starting address of the memory location also, memory block also where we are going to store the information or from which memory block we are going to transfer information to the output device, along with that it will give that amount of data to be transferred ok. So, these are the things that I am saying it is going to give data count. It is going to give the starting address where which memory blocks that we are going to use. So, these are the information that CPU is going to tell to the DMA controller.

Now, CPU carries out with other works. Now this now it needs to transfer some information, but now processor is giving that information; that means, it is delegate the job to the DMA controller. Now processor can carry out its own work. DMA controller will deal with the transfer, DMA controller sends interrupt when finished. Now deals with the after getting all those info required information now DMA controllers will deal with the transfer operation from your device to the memory or memory to the device once everything over then it will sends an interrupt to the processor saying that this finished, basically why it is required. Now at that particular point we are going to transfer the bus from the DMA controller to the processor now. So, that procedure can fetch information from main memory. So, these are the operations, generally we perform when we are going to do a DMA transfer.

(Refer Slide Time: 22:25)



So, what basically it does now basic difference over here is that DMA controller takes over the bus. So, now, system bus, basically data bus and address bus is used to connect the memory and along with that I/O devices also. So, when we are using DMA controller then this control of this particular bus will be taken over by DMA controller; that means, processor is no longer will use this particular bus, while DMA is transferring the information.

And so since DMA is taking order the control of the bus, now DMA is directly going to transfer a information from device to the memory or from memory to the devices. So, it is going to take over the bus then it is going to transfer the information. So, in that particular case it says that it is not an interrupt and we don't have or your CPU does not have switch context. Now this is the difference between your interrupt. In case of interrupt driven I/O what will happen.

In case of interrupt driven I/O what we are doing? We are running an interrupt service routine in the processor itself; that means, there is a context switch. So, basically processor is running a program it is in one particular context. This context is related to the program that was executed in the processor, but when interrupt comes, when we are going to give service to the interrupted devices; that means, we have to run the appropriate interrupt service routines. So, there is a sense of context in the processor, it is running one program, but it now processor is going to run another program in case of interrupt driven I/O.

So, for that we have to retain the processor status and how we are doing it? we are just transferring the status of the processor to the system stack and processor is now going to execute the interrupts service routine. So, there is a sense of context from the program being executed in the processor is now suspended, we are storing the context or the relevant information in the systems stack and now processor is going to execute the another program which is in that service routine, once the transfer is over then processor is going to restore its initial state; that means, again it is coming to the same context.

So, this is a context change ok. So, it is going to change its context from one program to the other program. But in case of DMA transfer there is no context change. Processor is executing one particular program it till going to execute the particular program or it will return state, because the data transfer operation has been delegated to that DMA controller. Now DMA controller is going to take care of transfer this information. So, processor is the same state, same context it can execute the same program provided the relevant information is available inside the processor.