Smart Card Tutorial

First Published in February 1995

Smart Cards From There to Here - part 1

There is something about the concept of a Smart Card that captures the imagination. If you talk to anybody in the business they invariably enthuse about this new technology yet its not really that new.

The Smart Card concept is based on the idea of embedding an integrated circuit chip within the ubiquitous plastic card that has become a part of our everyday life. The expression Smart Card was actually coined to describe a plastic card containing a microchip with processing capability. Today there are still more such cards containing just a memory chip which form the core of the telephone cards that have been so widely used in France and Germany. The more correct expression to cover all such cards is an integrated circuit card (ICC) as defined in the current ISO standards. However, in tracing the story of the technology we will be a little lax in our terminology to reflect the public's perception of these novel pieces of plastic.

One cannot help but wonder how it all happened. In fact the history of the Smart Card is really steeped in the development of chip technology over the last 40 years. Today we have small pieces of silicon just 5mm ? 5mm fabricated to provide a full microprocessor with various classes of memory. These chips are more powerful than the early personal computers that emerged in the early 70's. Some of these chips are capable of performing the latest public key cryptographic algorithms in fractions of a second that just 10 years ago would have seemed impossible.

Although the modern history of the Smart Card has seen its successful commercialisation promoted by Roland Moreno through his patents starting in 1974 there is much more to the story. Other pioneers such as Dethloff and Arimura whose work predated Moreno's patent each played an important part in the invention of the Smart Card and its applications. This story is both interesting and complicated and we shall describe the plot in a separate subsequent article.

One cannot help but ask, why put a chip in a plastic card. The engineers amongst us would probably point out how much easier it would be to fabricate a chip in a more solid module. It potentially would be more reliable because of the less stringent mechanical environment. There are in fact a large number of such devices fabricated in different form factors such as small plastic keys as would be synonymous with an access control device. The answer lies in the world of standards. A subject that to many people is both boring and unnecessary because the best product makers its own de-facto standard. Well of course this is partially true, International standards are invariably based on an existing product that has become the winner in its field. However the full commercial exploitation of a product needs interoperability without which the market becomes very fragmented. The competition between Betamax and VHS for the video tape recording standard was so significant that the loser, in this case Sony became very wary of ever again ending up with an unsupported product. Today we can see a very different approach in the development of digital video discs (DVD) where the major players are in close discussion to try to agree a common standard.

So on to the Smart Card, where are the standards? Well of course the form of the common plastic bank card that most of us have in our pockets is a widely accepted International standard. Its not just the card itself but the surrounding infrastructure through which the true interoperability can be achieved.

Today we are faced with a number of organisations promoting standards. In the field of Smart Cards we have various offerings from competing organisations in the definition of their products. Recently Europay, Mastercard and Visa (EMV) have published their specifications for Smart Cards as have Mondex (the National Westminster bank global payments electronic purse initiative). But here is the point they are specifications not standards. All these organisations are basing their products on the appropriate ISO standards but they define a more precise specification which describes exactly how their product works within the various options offered by the standards. These specifications allow interoperability in terms of the infrastructure (i.e the terminal interface) but may well be different at the application level. For example

the Mondex purse at the application level operates differently to the proposed EMV product but a terminal designed to meet ISO standards can easily operate with both products.

It is this common infrastructure which is the key factor to the success of a new technology and as we shall see ISO has developed a number of standards that are sufficient for the adoption of the Smart Card technology such that within a few years we shall probably all have one in our pocket. Marketeers might just like to add this up around the world, and please include China in your figures.

ISO uses the term, Integrated Circuit Card (ICC) to encompass all those devices where an integrated circuit is contained within an ISO ID1 identification card piece of plastic. The card is 85.6mm x 53.98mm x 0.76mm and is the same as the ubiquitous bank card with its magnetic stripe that is used as the payment instrument for numerous financial schemes.

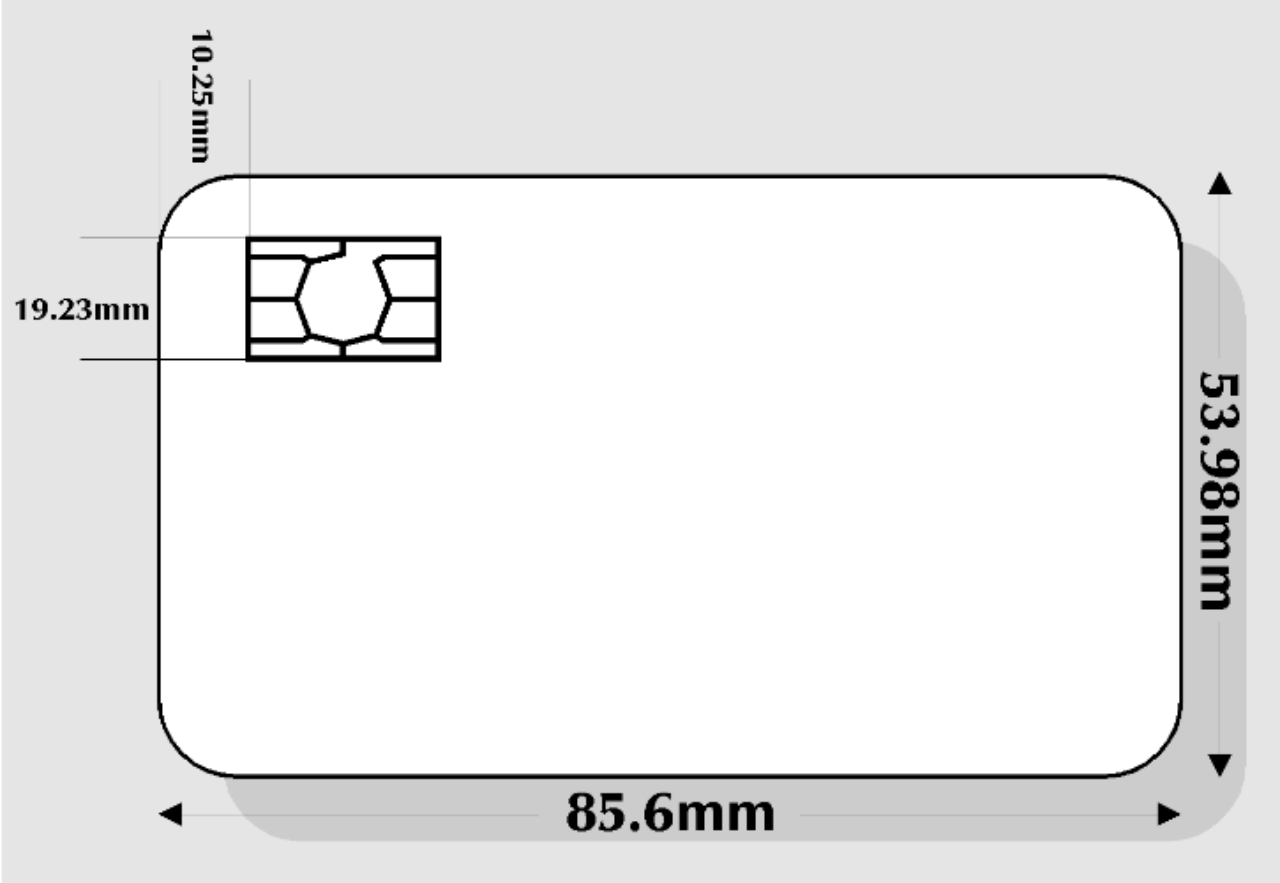


Figure 1: ISO ID 1 Card

Integrated Circuit Cards come in two forms, contact and contactless. The former is easy to identify because of its gold connector plate (fig 1). Although the ISO Standard (7816-2) defines eight contacts, only 6 are actually used to communicate with the outside world. The Contactless card may contain its own battery, particularly in the case of a "Super Smart Card" which has an integrated keyboard and LCD display. In general however the operating power is supplied to the contactless card electronics by an inductive loop using low frequency electronic magnetic radiation. The communications signal may be transmitted in a similar way or can use capacitive coupling or even an optical connection.

The Contact Card is the most commonly seen ICC to date largely because of its use in France and now other parts of Europe as a telephone prepayment card. Most contact cards contain a simple integrated circuit although various experiments have taken place using two chips. The chip itself varies considerably between

different manufacturers and for a whole gambit of applications. Let us consider first the purpose for the 6 contacts used by the ICC (fig 2).

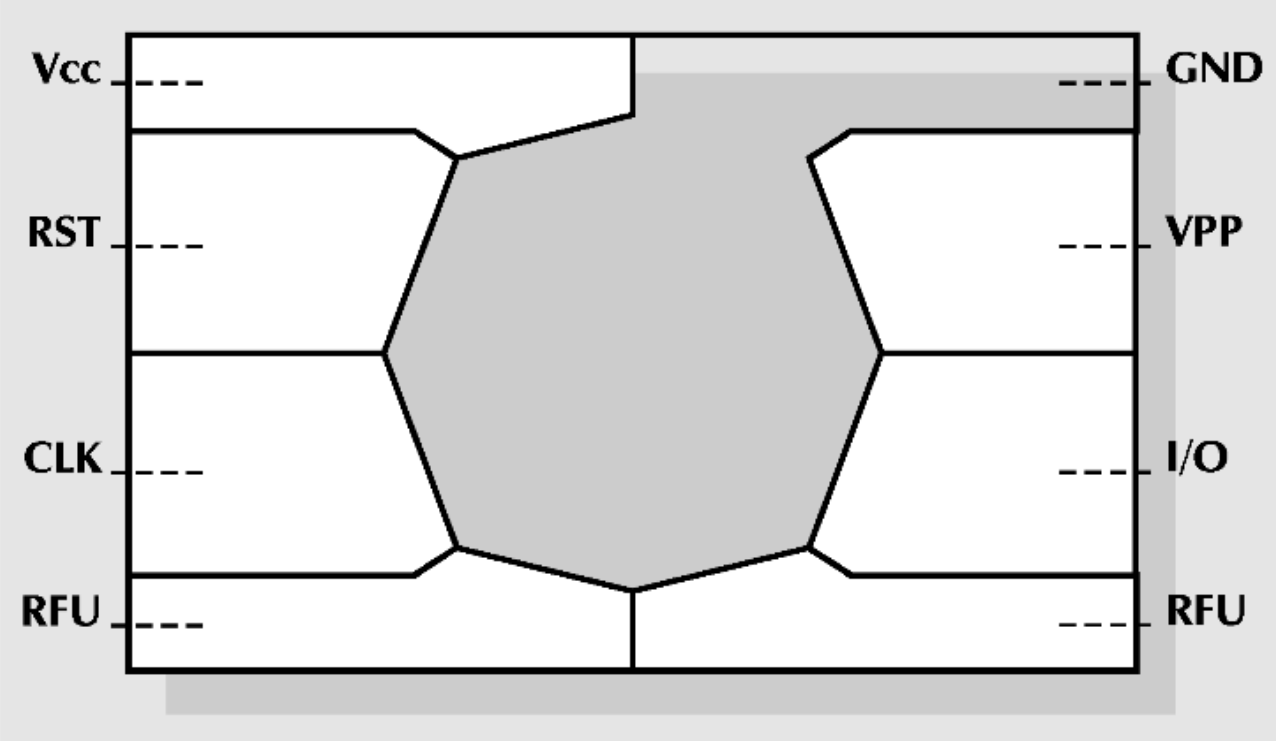


Figure 2: ISO 7816-2 Connector

Vcc is the supply voltage that drives the chips and is generally 5 volts. It should be noted however that in the future we are likely to see a move towards 3 volts taking advantage of advanced semiconductor technology and allowing much lower current levels to be consumed by the integrated circuit. Vss is the substrate or ground reference voltage against which the Vcc potential is measured. Reset is the signal line that is used to initiate the state of the integrated circuit after power on. This is in itself an integral and complex process that we shall describe later in more detail.

The clock signal is used to drive the logic of the IC and is also used as the reference for the serial communications link. There are two commonly used clock speeds, 3.5795 MHz and 4.9152 MHz. The lower speed is most commonly used to date in Europe but this may change in the future. One may be tempted to ask why these strange frequencies were chosen, why not just a straight 5 MHZ. The reason lies in the availability of cheap crystals used in the television world. For example the American NTSC colour subcarrier frequency is exactly 3.579545 MHz. The Vpp connector is used for the high voltage signal that is necessary to program the EPROM memory. Last, but by no means least is the serial input/output (SIO) connector. This is the signal line by which the chip receives commands and interchanges data with the outside world. This is also a fairly complex operation and will be the subject of a more detailed discussion where symbols such as T=0 and T=1 will be fully explained.

So what does the chip contain, well the primary use of the IC card is for the portable storage and retrieval of data. Hence the fundamental component of the IC is a memory module. The following list represents the more commonly used memory types,

- ROM Read only memory (mask ROM)
- PROM Programmable read only memory
- EPROM Erasable programmable ROM
- EEPROM Electrically erasable PROM
- RAM Random access memory

A particular chip may have one or more of these memory types. These memory types have particular characteristics that control their method of use. The ROM type of memory is fixed and can not be changed once manufactured by the semiconductor company. This is a low cost memory, in that, it occupies minimum space on the silicon substrate. The use of the silicon is often referred to as real estate because clearly one wants to get as much as possible into the smallest possible space. The snag however is that it cannot be changed and takes several months to be produced by the semiconductor company. There is also effectively a minimum order quantity in order to achieve this low cost.

In order of increasing real estate the PROM comes next. This memory is programmable by the user through the use of fusible links. However high voltage and currents are required for the programming cycle and such devices are not normally used in Integrated Circuit Cards. The EPROM has been widely used in the past but the name for this application is something of a misnomer. Whilst the memory is erasable, by means of ultra violet light, the necessary quartz window is never available in the ICC and the memory is really used in one time programmable mode (OTP). Getting pretty heavy in real estate terms is the EEPROM. This memory is indeed erasable by the user and can be rewritten many times (between 10,000 and 1,000,000 in a typical implementation) All of these memories described so far are non volatile. In other words when the power is removed they still retain their contents. The random access memory (RAM) is a different kettle of fish, this is volatile memory and as soon as the power is removed the data content is lost.

In order to pursue our studies further we must note that the cost of the IC at saturation (i.e when development costs have been recouped) is proportional to the square area of silicon used (assuming constant yield). The ISO connector is so designed to constrain the silicon die size to about $25mm^2$ (although it is possible to handle $35mm^2$ or more). However the important point is more concerned with reliability where clearly the larger die will be more prone to mechanical fracture. There is another bi-product that we will consider later where the cost of testing and personalisation are considerably altered by the complexity of the particular chip. It is clear however that we should attempt to minimise the contents of the chip on both cost and reliability grounds commensurate with the particular application.

Well of course you cannot have something for nothing and although a telephone card may operate with a little EEPROM memory (128 - 512 bytes) and the memory control logic, more sophisticated applications will demand ROM, EEPROM, RAM and a CPU (Central Processing Unit) to achieve the necessary business. It is the addition of the CPU or micro-controller that really leads to the term "Smart" although as mentioned previously we will not be rigorous in our use of the term.

The control logic should not be overlooked as this is necessary not only for communication protocols but also to offer some protection of the memory against fraudulent use. The ICC is probably the security man's dream because unlike most electronic storage and processing devices it has security intrinsically built in. The ICC really does provide a tamper resistant domain that is difficult to match with the somewhat larger security boxes that handle cryptographic processes.

So now we can differentiate the different types of ICC by their content,

- Memory only
- Memory with security logic
- Memory with CPU

The security logic can be used to control access to the memory for authorised use only. This is usually accomplished by some form of access code which may be quite large (64 bits or more). Clearly the use of EEPROM memory must be strictly controlled where fraudsters can obtain a financial advantage by unauthorised use. This applies as much to telephone cards as applications using ICCs for cryptographic key carriers. The security advantage of the CPU device is of course more significant because the CPU is capable of implementing cryptographic algorithms in its own right, but we will discuss this in more detail in due course.

In the Smart Card world the term application is widely used to describe the software or programs that the IC implements. In the simplest case the application may be just a file manager for organising the storage and retrieval of data. Such an application may be totally implemented in the logic of the chip. Similarly the chip must contain the communications logic by which it accepts commands from the card acceptance device (CAD) and through which it receives and transmits the application data. The ICC which contains a CPU can handle more sophisticated applications and even multi applications since the CPU is also capable of processing the data and taking decisions upon the various actions that may be invoked. The subject of multi-applications and particularly the implementation of security segregation is another subject for more detailed discussion in subsequent parts.

by Dr. David B Everett
Next Month: Part 2 - The Making of a Chip

Smart Card Tutorial

First Published in March 1995

From there to here - part 2

The making of a chip

The chip is what it's all really about, after all the plastic card is just the carrier for the embedded chip fabricated on its associated micro module. We cannot really appreciate the issues surrounding the use of a Smart Card without some understanding of the underlying technology. Its a bit like driving a car without knowing what's under the bonnet. You can certainly get from A to B but you don't really have the basis of the knowledge necessary to analyse the economics, performance, reliability and security.

The technology of the chip is one of the wonders of modern science. The advances made over the last twenty years are probably unsurpassed by any other technical invention. The key to the advances in this technology are geared to size or how to put twice as much in the same space as last year. In 1971 we saw the introduction from Intel of the 4004 the worlds first microprocessor. This chip was made up of 2300 transistors on a piece of silicon about 3mm by 4mm. Today this same company is producing the Pentium microprocessor about ten times the size but containing over 3 million transistors. Those of us brought up in the era of the valve will marvel at the thought of interconnecting 3 millions valves to make something that works reliably for many years without failure. It would have been an impossible dream.

The transistor at the centre of this technology is an apparently simple device but readers are warned that the subject of semiconductor physics is not a matter to be entered into lightly. We can cheat however and jump straight into the meat of the matter. In figure 1 we show an MOS transistor. The MOS comes from Metal, Oxide, Silicon which makes up the sandwich of the transistor. The conducting gate material used to be made of metal but today it is generally made from doped polycrystalline silicon which is a reasonable conductor. The insulating layer is silicon dioxide which is an extremely good insulator. The final layer in the sandwich is the lightly doped silicon semiconductor substrate. One of the most useful comments to make here is how the different forms of silicon can be used to create anything from a perfect insulator (Silicon dioxide) to a conducting track (doped polysilicon) as typically used to make the gate connection. The doping of the silicon is fundamental to altering its conductivity. Two types of doping are used, one produces n - type silicon which is rich in electrons (negative charge) and the other p - type which is rich in holes (electron duality to produce positive charge).

In figure 1 we can see the two types of MOS transistor the n - channel and the p - channel. In a CMOS device which is the form of most Smart Card chips today both channel types are fabricated in a single chip. The input to the transistor is the gate which by means of an applied voltage controls the current flowing through the output terminals, the source and drain. These terminals are interchangeable but by convention the source is the common reference terminal to the input and output current while the drain terminal is connected to the output load. The source current is equal to the drain current since there is negligible gate current and as shown in figure 2 by convention it is in the opposite direction to the flow of electrons.

The operation of the transistor is controlled by applying a potential field across the gate insulator. Assuming the applied voltage is positive then as shown in figure 2 a negative charge is induced on the semiconductor which repels the holes in the p - type substrate. This forms what is called a depletion region on the surface of the substrate. As the field in the gate region is increased electrons are drawn into the channel region under the gate from the heavily doped n - type source and drain regions. When this electron density is higher than the hole density then a surface inversion channel is formed. At this point the channel becomes resistive between the source and drain. As the gate voltage is increased further current flows across the channel and the transistor is switched on. This basic transistor operation is the core of the building blocks for the logic in the microprocessor by which various gates (AND, OR etc) can be constructed to implement the overall functionality of the device.

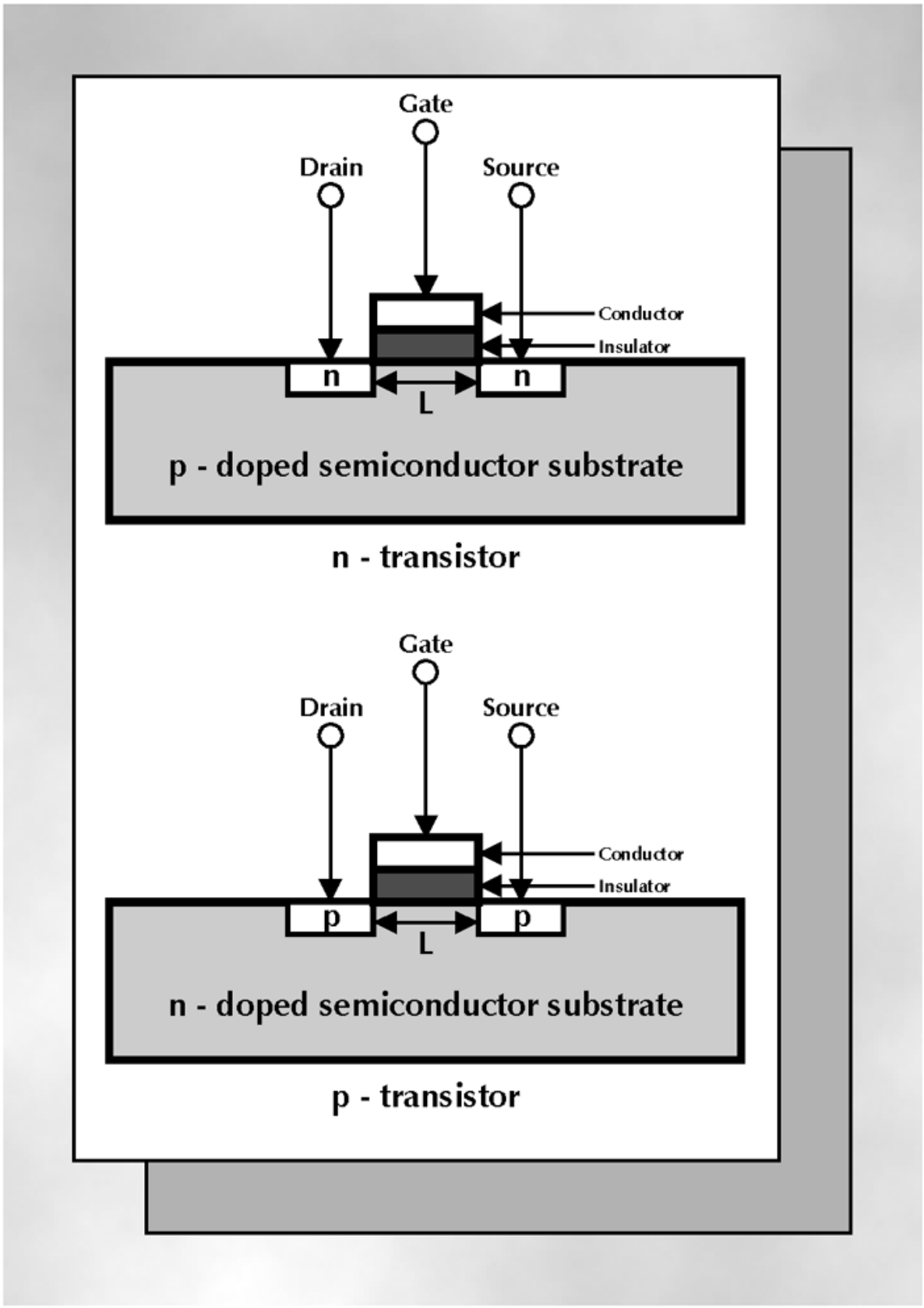


Figure 1: The MOS Transistor

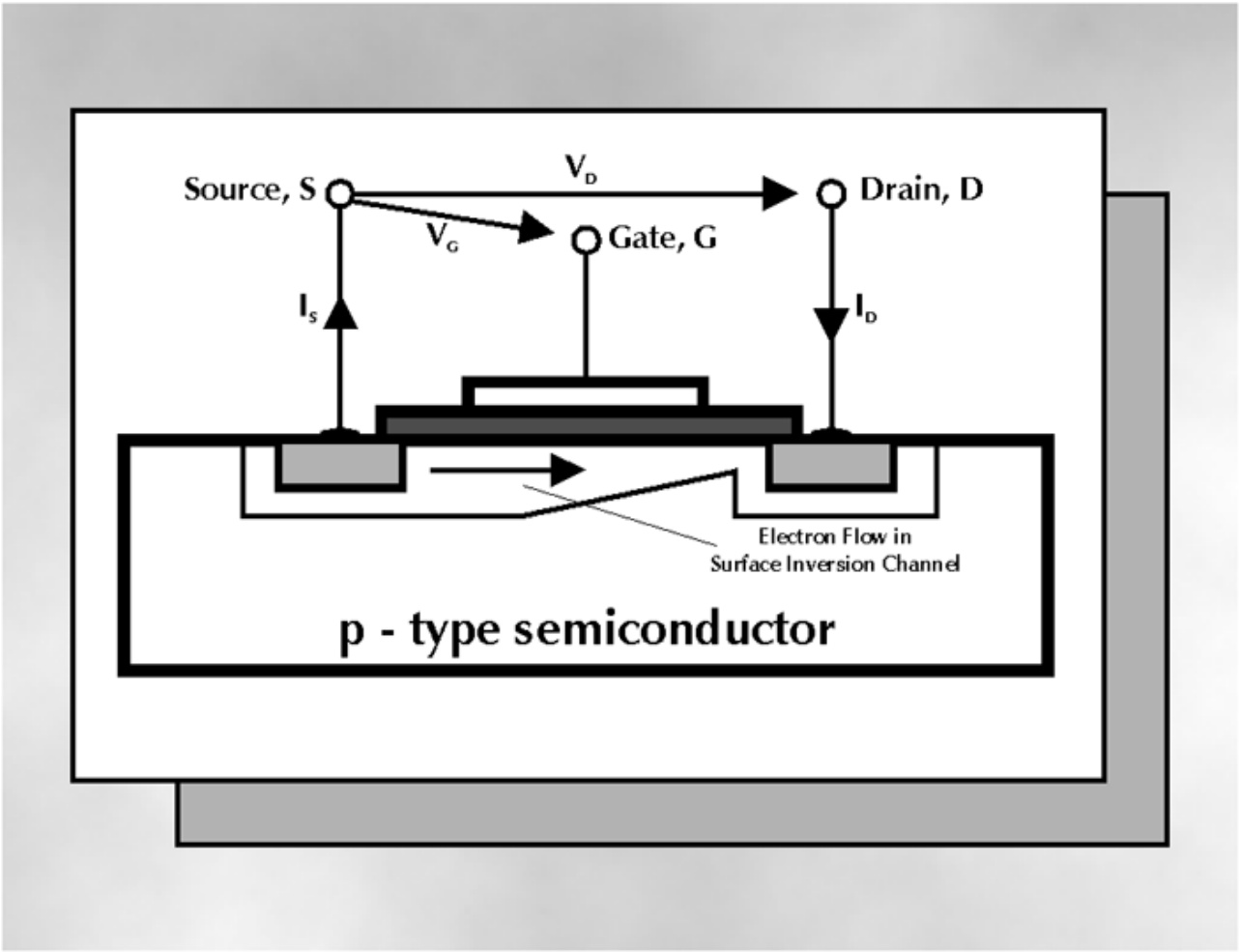


Figure 2: Operation of the MOS Transistor

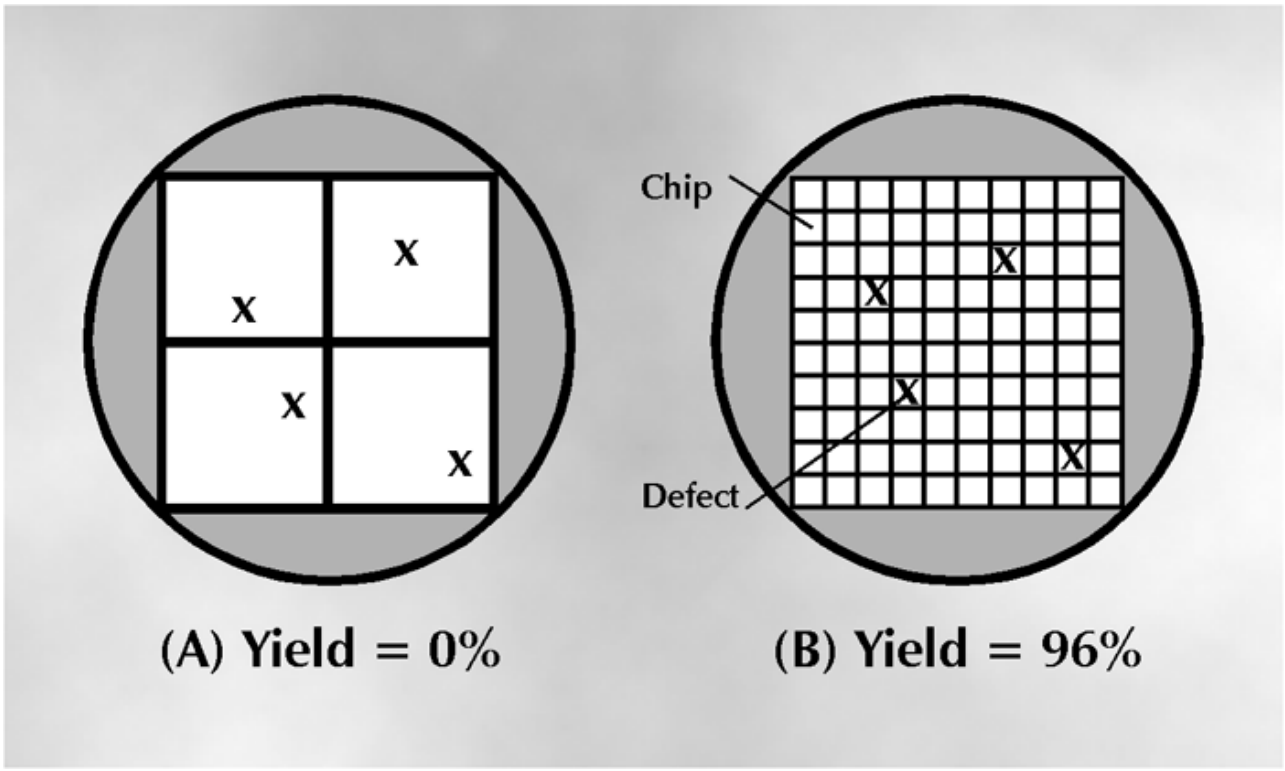


Figure 3: Yield as a Function of Defects

Before moving on to the actual fabrication of the chip on the silicon wafer we need to examine some of the concepts of size which are fundamental to our understanding of the cost of the final chip. First of all lets look at the size of the gate region shown in figure 1 where W and L represent the width and length of the gate. This is the smallest feature size on the chip and is still limited (For a few years yet) by the process technology used to make the wafer. When you hear 1 micron (I millionth of a metre = 10^{-6} m) technology what is being referred to is the minimum feature size that can be fabricated. The total transistor is of course much bigger than the gate region and has to allow for the source and drain areas as well as the necessary interconnections. As a first approximation the average space per a transistor is about 100 times the gate size. The Pentium microprocessor is about 100mm^2 and is fabricated with a 0.6 micron process (moving to 0.4 micron in 1995).

Thus the minimum gate size is 0.6 μmx 0.6μm. Taking the total area of the chip divided by the 3 million transistors we end up with an average area per transistor of 33μm^2. These are pretty small figures, just for comparison the human hair is typically 300 microns in diameter. By the way that gate insulator, the silicon dioxide, well in some cases that may only 100 Angstroms thick ($1A^0$ = 1 ten billionths of a metre = 10^{-10} m = average width of an atom).

When you get these dimensions in perspective it is readily apparent that a few defects on the wafer will significantly alter the yield (Figure 3) and this is one of the reasons why the size of the chip is so important to the final cost of the chip since the wafer processing cost is relatively constant at a few hundred dollars per wafer.

David B Everett

Next month - Making the chip - continued.

Smart Card Tutorial

First Published in April 1995

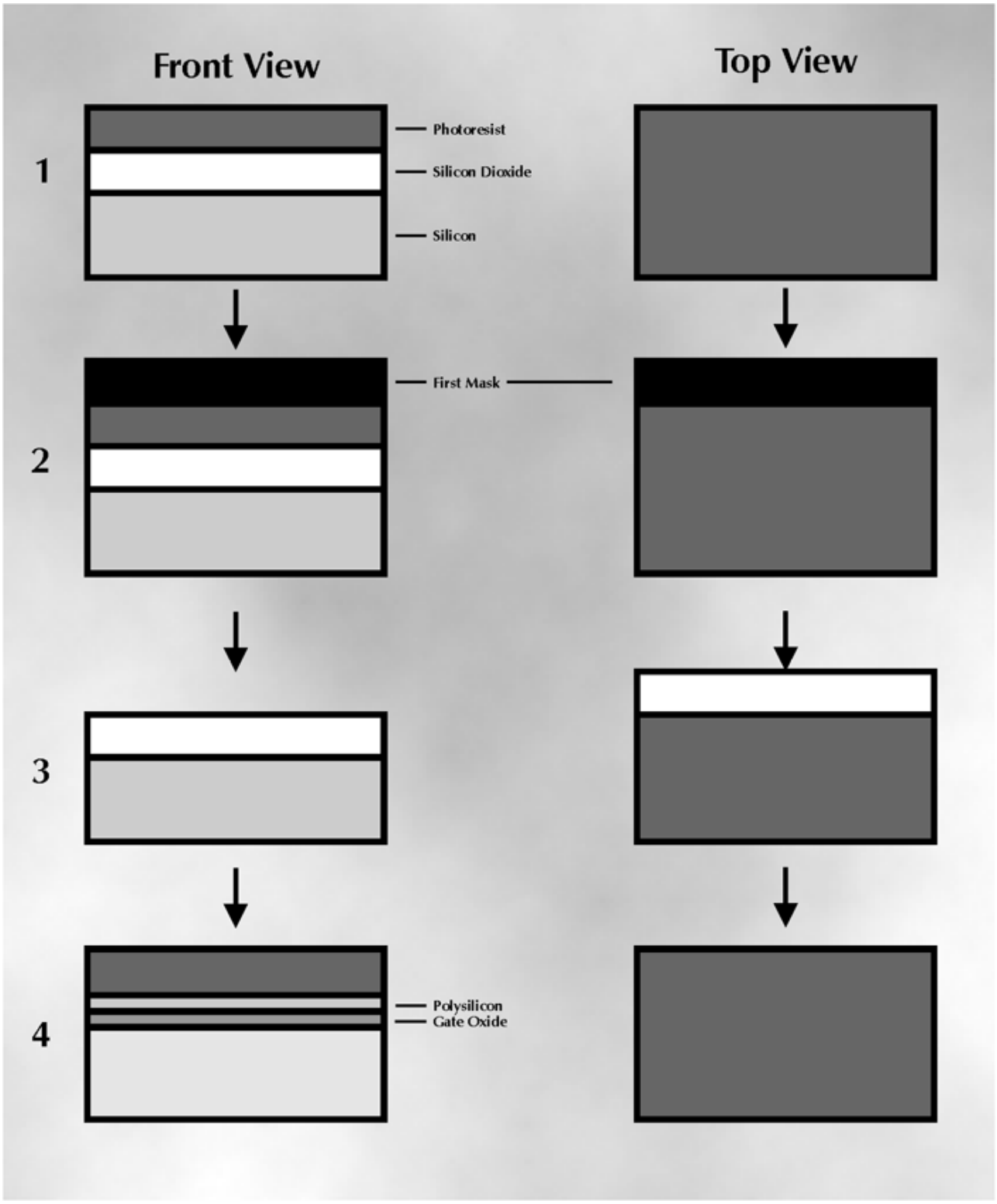
From There to Here - Part 3.

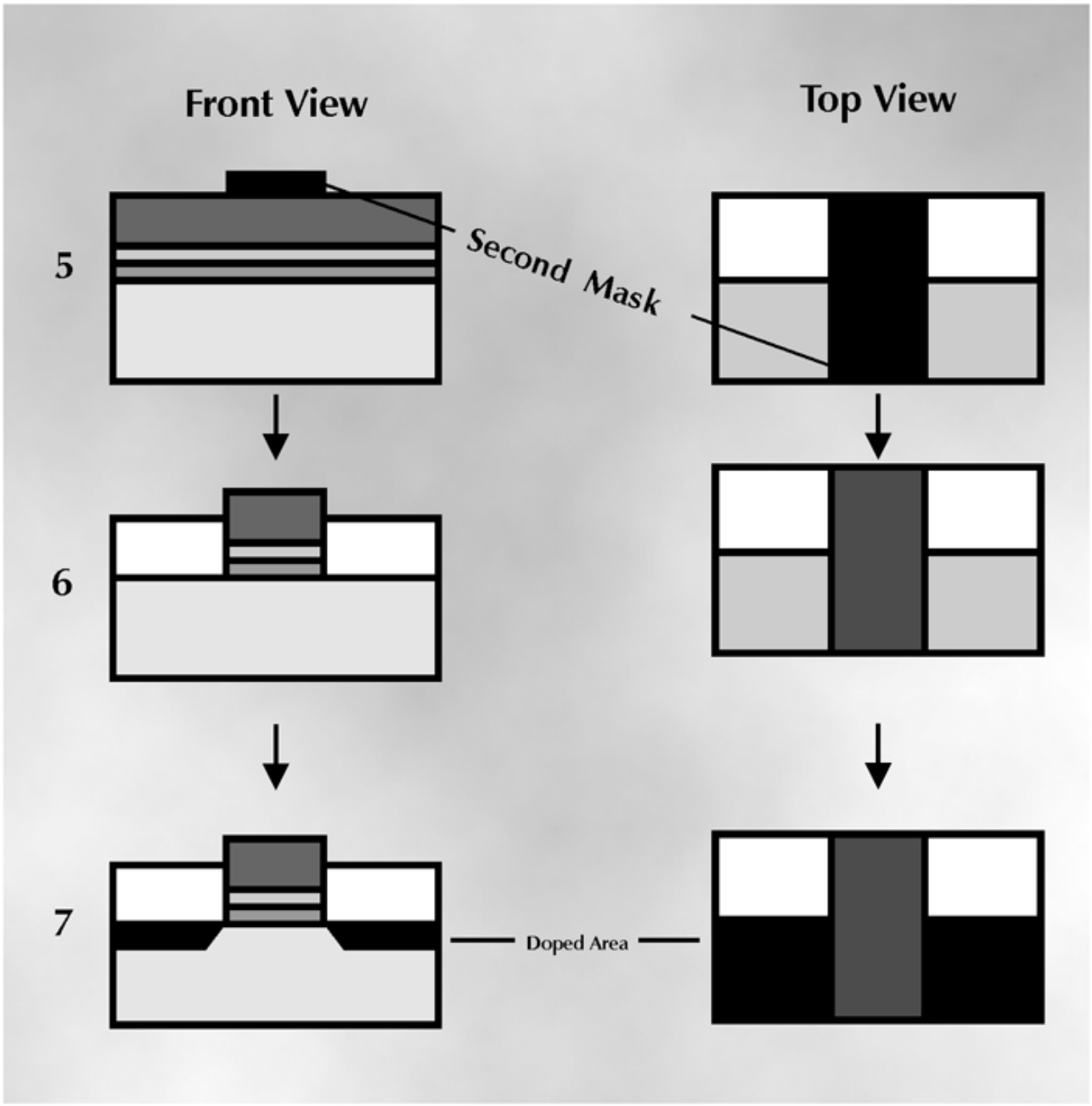
The making of a chip continued.

The MOS transistor is the funamental building block of the Smart Card chip. An understanding of the principles underlying its fabrication are necessary in order to fully appreciate the potential of the chip from both a performance, cost and security point of view. As an example the drawings show the major steps in the manufacture of an n-channel MOS transistor which is formed in a p-type substrate.

The source and drain regions are formed by selectively converting shallow regions at the surface to n-type material. The insulator, silicon dioxide is formed on the surface to create the gate insulator and also to separate one device from another. Aluminium metal is used to make the connections to the source and drain whilst conductive polysilion is used to make the connection to the gate.

A number of processes are used during the course of the chip manufacture:


Front View
Top View
1
Photoresist
Silicon Dioxide
Silicon
2
First Mask
3
4
Polysilicon
Gate Oxide



Oxidation

Silicon dioxide can be formed on the surface of the silicon by heating the wafer to a high temperature (1000-1200^{0}C) in the presence of oxygen.

Photolithography

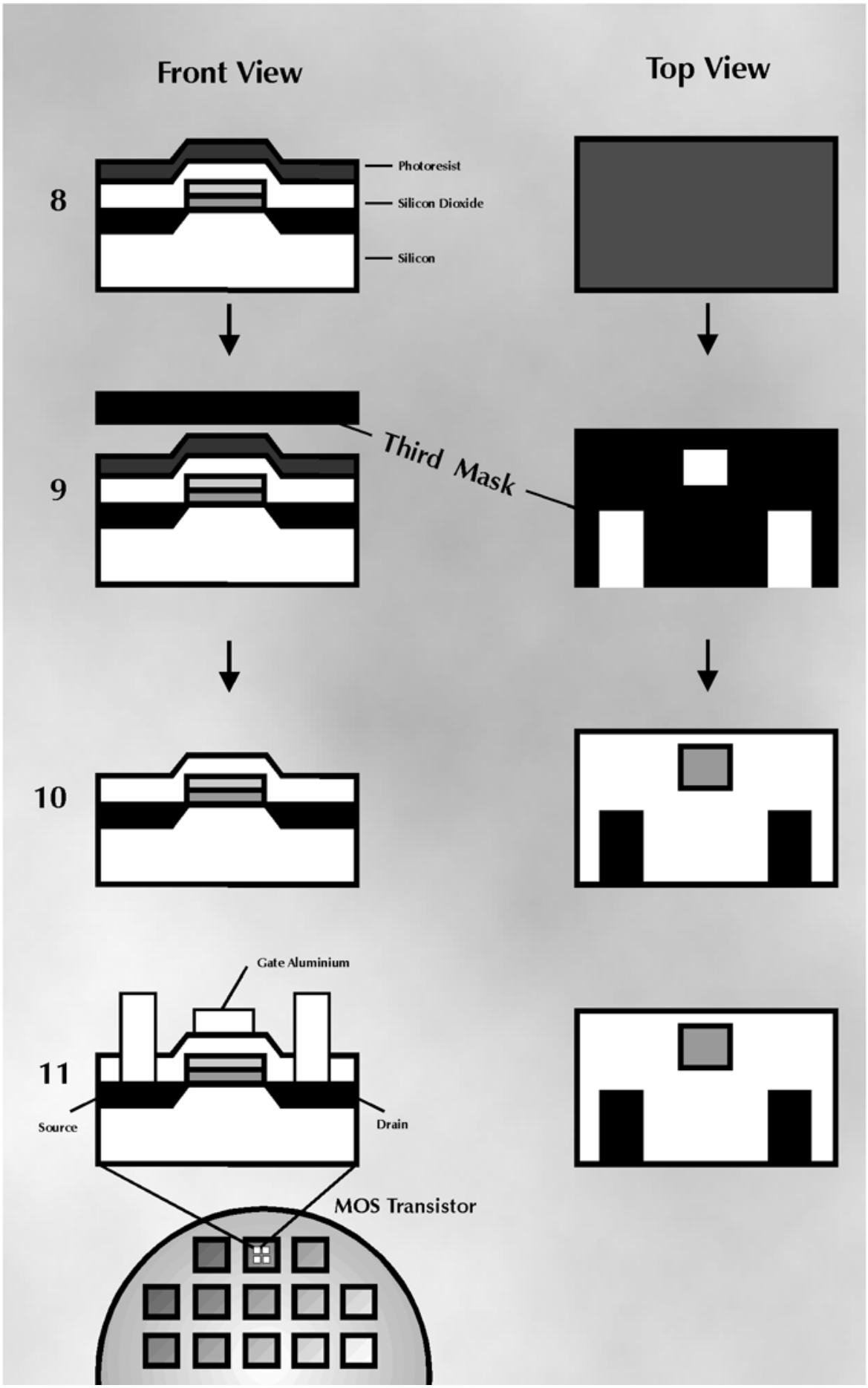
This is the basic mask process where the surface is covered in a photo resist which becomes soluble when exposed to ultra violet light. The mask that is interposed between the wafer and the source of light is made by depositing the relevent pattern in a light absorbent material on a glass plate.

Etching

Etching with acids or a gas plasma is used to cut patterns in the masking material. Silicon dioxide is the principle masking material but polysilicon and silicon nitride are amongst the materials that may be used to mask areas against ion implantation or diffusion.

Ion implantation

This is a process where by the wafer is bombarded by high energy atoms generated in a high voltage particle accelerator. The atoms are donor type (in our example) to alter the characteristics of the bombarded area to form the shallow n-type layers of the source and drain.


Front View
Top View
8
Photoresist
Silicon Dioxide
Silicon
9
Third Mask
10
11
Gate Aluminium
Source
Drain
MOS Transistor

Diffusion

This is an alternate process to Ion implantation for forming the shallow n or p type regions. The donor or acceptor impurities are diffused into the surface by raising the wafer to a high temperature (>1200^{0}C).

Evaporation

Metal films that make up the conductive interconnections are deposited by evaporation when the metal is heated to its melting point in a vacuum.

Chemical Vapour Deposition (CVD)

Thin films of silicon dioxide and polysilicon can be deposited out of a gaseous mixture onto the surface of the wafer. This process is typically applied at high temperatures and low pressure.

Epitaxy

CVD can be used to deposit silicon onto the surface of the wafer. A single crystal silicon layer may effectively be grown on the surface of the wafer by this process.

David B Everett.

Next month - More into the economics of chip manufacture.

Smart Card Tutorial

First Published in May 1995

From there to here - Part 4

How much does the chip cost?

You might argue that this is starting at the wrong end. How can you know what it will cost before you have determined what you want. The main reason for tackling this subject so early is that the Smart card is one component in a system. It is the over all cost of the system that matters and if we can understand the cost of the Smart Card in terms of what you get for your money then it enables us to apportion the various tasks in a system to the optimum components.

At the very least the Smart Card is a portable carrier of data. It is difficult to justify the use of a Smart Card without this attribute. The role of a portable processor in such an expensive format has limited value. So we could start off by defining the Smart Card chip as a memory device with the necessary interface logic and the option of a data processing unit. If we assume for the moment that our Smart Card does not contain a battery (if it did the battery would be most expensive component) then clearly the memory needs to be non volatile. In other words the data must be preserved during power off. This means that the integrity of the data must be assured (1's must not change to 0's or vice versa). And of course we can add a further condition which says that during the course of normal usage only. Hitting a Smart Card chip with a hammer seems just a little unfair. For the purpose of Smart Cards we propose a definition of normal environment as being that in which a human being (the card carrier) might reasonably expect to survive. In fact the Smart Card chip will do better than the human being in almost every case. Survival of the chip is determined primary by mechanical strength and here the chip must not be subject to abuse such as the hammer or the pressure of the ball point pen used by some French consumers to force their bank Smart Cards into the fall back mode.

So if non-volatile memory is the name of our game what are the choices,

- ROM Read Only Memory
- EPROM Electrical Programmable ROM
- EEPROM Electrical Erasable PROM
- FLASH Flash Erasable Memory
- FRAM Ferro Electric Random Access Memory

Previously we have stressed that the cost of the chip (at maturity) is proportional to the area of silicon used. In looking at these different memory technologies the number of transistors used per memory cell is the primary factor in calculating the area of silicon used.

A ROM cell is a transistor that is either present or absent. The term mask ROM is often used because the presence or other wise of the transistor is a function of the mask used in the manufacture of the chip. In other words the chip is made either with or without the transistor in that particular memory cell location. In terms of silicon area this is about as efficient as you can get but quite clearly you can only read the state of the memory cell, it can not be subsequently changed after manufacture. Mask ROM is only normally used when the state of the contents of the ROM memory represent a stable situation. Thus a Smart Card chip operating system would be looked on as relatively constant and would justify a long production run and the initial set up costs in making the necessary manufacturing masks. Clearly you can not use the mask ROM memory for storing any form of variable application data.

The EPROM memory cell is again constructed by a single transistor per cell but here the device has two gates, a floating (storage) gate and a control gate. These two gates are fabricated one above the other over the channel separated by an insulating layer. The EPROM memory is normally erased by exposure to ultra violet light. This is not practical for a Smart Card application so the use of an EPROM memory really relates to a one time programmable memory cell. The memory cells are programmed by applying a high voltage

(12-25 volts) to both the drain and control gate whilst holding the source at ground. This causes electrons to flow through the insulating oxide from the substrate thereby charging up the gate. This process is called hot electron injection. When the floating gate is changed its turn on threshold voltage is increased and therefore when the transistor memory cell is normally selected the transistor no longer conducts.

The use of EPROM memory cells was most commonly seen in early Smart Card chips. Many applications can be operated with the concept of memory that can be written (programmed) once and subsequently cannot be changed. For instance telephone cards could use such a memory where the cells are programmed to indicated the usage of telephone units. There are problems in using such memory technology however which is why they are not commonly seen today. It was mentioned previously that a high voltage is required to program the memory cells. In terms of the ISO 7816 standard this is referred to as V_{PP}. This programing voltage was supplied by the terminal to the chip through one of the contacts on the Smart Card connector plate. Two problems arose, in the first instance some of the early terminals miscalculated the correct voltage (V_{PP}) to apply to the chip and the subsequent over voltage resulted in the rather permanent expiry of the chip. The opposite situation was observed by the hacking population who noticed that if you prohibit the V_{PP} voltage from reaching the chip then the memory write (or program) could easily be inhibited. This was a useful way of making free phone calls or even receiving free subscription satellite television viewing.

The electrically erasable read only memory (EEPROM), sometimes called E squared is the main stay of current Smart Card chips. This memory can be electrically programmed and erased, ie it is true read and write non-volatile memory.

The problem with EEPROM is its size, each memory cell consists of 2 transistors, a select transistor and a storage transistor. Some times the EEPROM cell is referred to as read mostly since it has a limited endurance. Although it can be read a practically infinite number of times the erase and write cycles are technology limited, typically to about 100 thousand cycles. It should also be noted that a high voltage is required to erase and write the memory cell but here the high voltage is usually generated on chip using a charge pump. For EEPROM memory a typical erase/write cycle for a memory cell is 10mS. In practice this is long time, for comparison the random access memory (RAM) of a micro controller is probable capable of 100nS or less (ie 100 thousand times faster).

The FLASH memory technology is relatively new for Smart Card chips and uses similar technology to the EPROM in that there is only one transistor per memory cell. It is also limited in its number of erase/write cycles. The electrical erasure is either for the entire memory array or for blocks of the array. This makes the memory very suitable for constant data such as application program modules but less suitable for dynamic application data.

The FRAM or ferro electric random access memory is the newest technology which has been developed by Ramtron. The ferro electric effect is the ability of a material to retain an electric polarisation in an absence of an applied electric field. Current memory cells use two transistors and two ferro electric capacitors. However new devices are expected to be available that only use one transistor and one ferro electric capacitor per memory cell. The FRAM has a 400nS read/write cycle time and has an endurance of 10 million cycles. It should be noted however that the endurance includes standard read cycles as well as write cycles. FRAM technology is not yet available in any chips designed for Smart Card applications.

If the Smart Card chip includes a micro controller then it will be necessary to include some random access memory (RAM). There are two types of RAM, dynamic RAM (DRAM) and static RAM (SRAM). A DRAM memory cell is made up of one transistor and one storage capacitor whereas static RAM uses six transistors. Unfortunately the DRAM memory which is widely used in computers is not really viable for Smart Card micro controllers because the memory needs to be constantly refreshed to allow for leakage from the storage capacitor. For Smart Card chips the SRAM is the biggest user of the valuable silicon real estate, 256 bytes of RAM may take up to 1/5 or more of the whole chip. In fig. 1 we show an example floor plan of a Smart Card micro controller chip showing the relative area taken by each component of the chip. In a typical example the EEPROM might be 8K bytes, the ROM 12K bytes and the RAM 256 bytes. In the example shown the total chip is $25mm^2$.

The cost for such a chip at maturity and in volume would be about $2. By comparison a simple EEPROM memory chip (of say 512 bits) with some security logic could be fabricated in 2mm2 of silicon with a chip cost of nearer 20 cents.

The examples here are based on the use of 1 micron technology where the smallest feature size is 1 micron. As the technology advances the smaller feature size will allow more transistors to be packed in the same area. Today 0.35 micron technology is routinely being used for DRAM memory devices and laboratory fabrication units are operating at about 0.1 micron. We have already mentioned that the cost of processing the wafer is relatively constant at a few hundred dollars. However the cost of these fabrication lines is increasing enormously. Today 500M dollars is not unusual and the cost is projected to reach 2 Billion dollars by the end of the decade. The problem here is the investment in these enormous capital costs. This is a game for big players where you need to fund the cost of the new generation from the current generation. Its a bit like the housing game in an inflating market, coming in late is a very expensive proposition. Today there are predominantly six companies producing microcontroller chips for Smart Card applications,

- Hitachi
- Motorola
- Oki
- Philips
- SGS Thomson
- Siemens

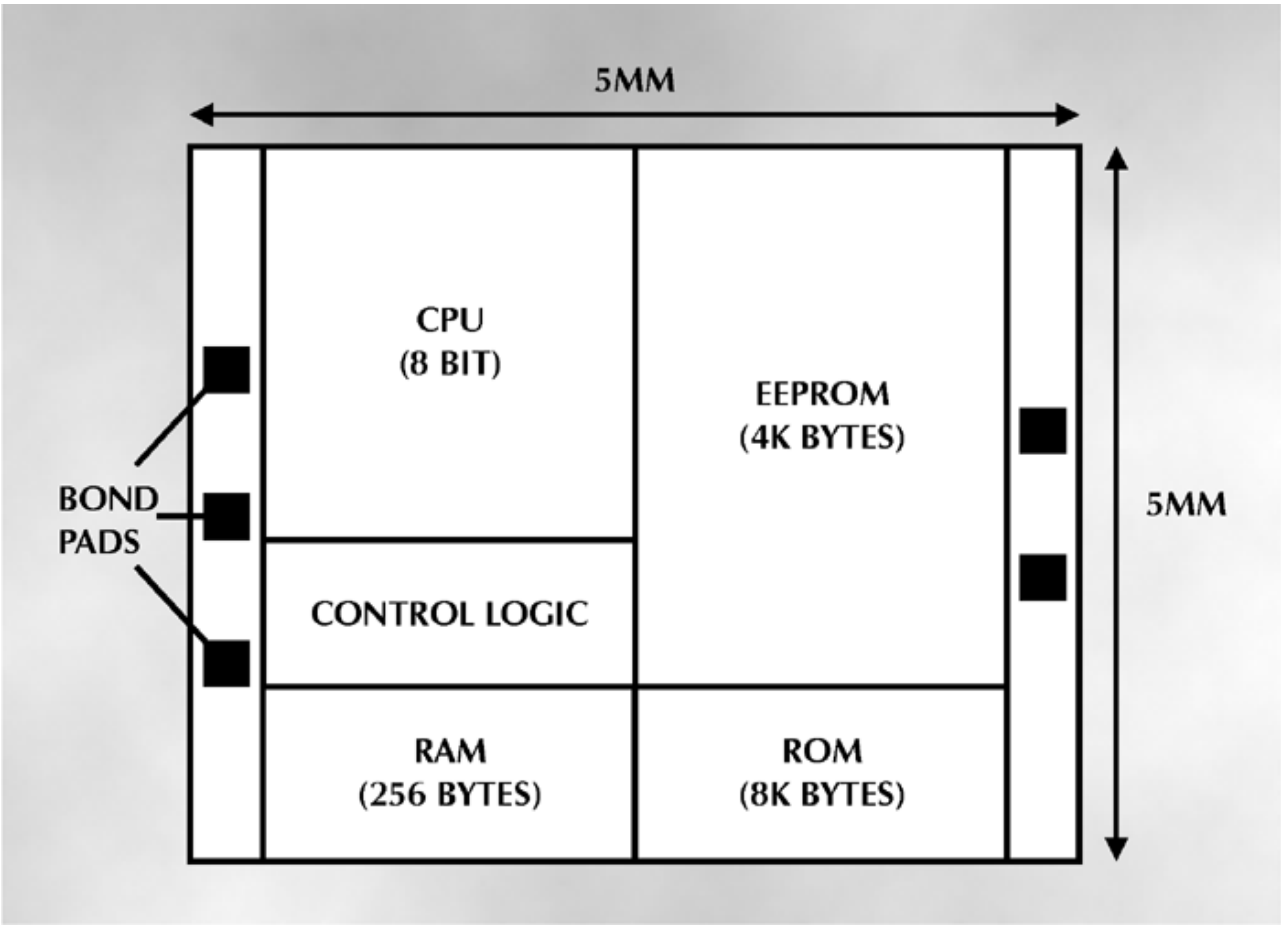


Figure 1: Typical Silicon Floor Plan for a Microcontroller Chip

There are also one or two others players in the wings that may well move into the gamebut what is clear is that there will never be that many companies and that the list above is more likely to get shorter. In any event the lowest cost will only come by taking advantage of the new technology and here we have one of our first dilemas. All Smart Card chips today operate at 5 volts, chips developed in the new sub-micron technology at say 0.65 microns or less would operate at 3 volts or lower. You cannot easily support a 5 volt infrastructure into the next century. This is currently occupying the minds of experts developing the ISO standards for Smart Cards. It is important not to assume that these new chips will be capable of withstanding 5 volts without damage.

Well back to our initial question, ? what does the chip cost ?. What we have tried to show here is that the price of the chip is largely a function of the size and type of memory. This is the area that the system designer should concentrate upon. The higher security applications will almost certainly require a CPU where the memory can be managed in a more controlled and secure way. General purpose multi application type chips by their very nature will require a large memory and therefor the higher cost. The use of FLASH memory is also consistent with this type of application and may become more commonplace in Smart Card chips of the future.

David Everett
Next month: Part 5 The total Smart Card manufacturing process.

Smart Card Tutorial

First Published in June 1995

From There To Here Part - 5

How the IC card is made

The manufacture of a smart card involves a large number of processes of which the embedding of the chip into the plastic card is key in achieving an overall quality product. This latter process is usually referred to as card fabrication. The whole operation starts with the application requirements specification. From the requirements individual specifications can be prepared for the chip, card, mask ROM software and the application software. The ROM software is provided to the semiconductor supplier who manufactures the chips. The card fabricator embeds the chip in the plastic card. It is also quite normal for the fabricator to load the application software and personalisation data. Security is a fundamental aspect in the manufacture of a smart card and is intrinsic to the total process. However we will consider security separately in subsequent articles in this series. We will look at each of the stages in the manufacture of the smart card as shown in fig. 1.

Chip specification

There are a number of factors to be decided in the specification of the integrated circuit for the smart card. For the purpose of this discussion we will consider a CPU based card although the manufacture of a memory card is substantially a subset of that described here. The key parameters for the chip specification are as follows,

- Microcontroller type(eg 6805,8051)
- Mask ROM size
- RAM size
- Non volatile memory type (eg EPROM, EEPROM)
- Non volatile memory size
- Clock speed (external, and optionally internal)
- Electrical parameters (voltage and current)
- Communications parameters (asynchronous, synchronous, byte, block)
- Reset mechanism
- Sleep mode (low current standby operation)
- Co-processor (eg public key cryptography)

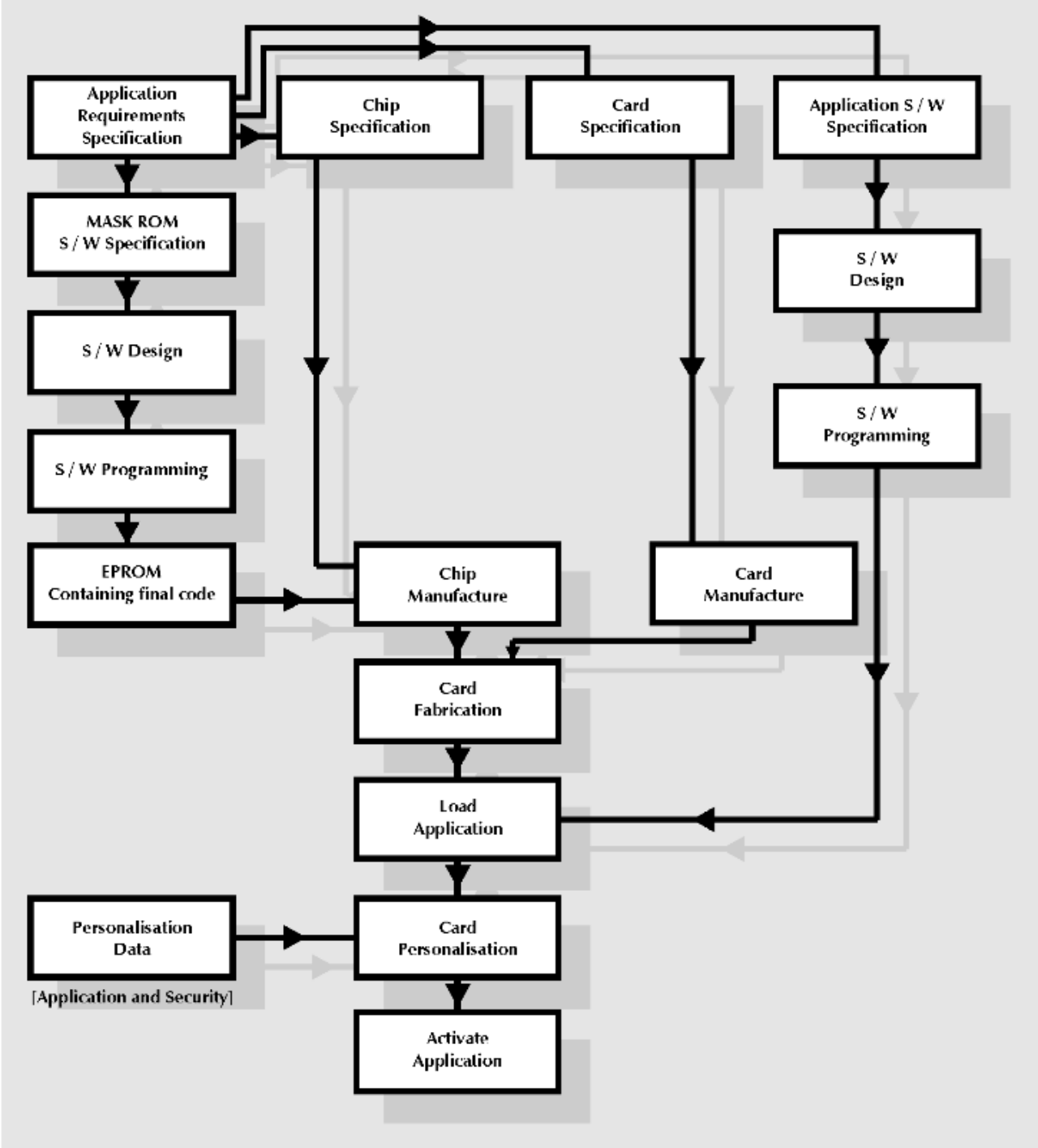


Fig 1 The Complete Process

In practice the semiconductor manufacturers have a range of products for which the above parameters are pre-defined. The task of the designer is therefore concerned with choosing the appropriate product for the particular application. As mentioned previously security may be an important issue for the application and accordingly there may be extra requirements on the physical and logical security offered by the particular chip.

Conformance to ISO standards is also likely to be a requirement and in this area ISO 7816 - 3 (Electronic signals and transmission protocols) is the principle standard to be considered. It should be noted however that ISO is currently producing a revision for part 3 of the ISO7816 standard. This revision is likely to alter some of the electrical characteristics of the chip. In particular there is the need to allow for low voltage (3V

or less) chips and lower current supplies as required in portable battery equipment. ETSI (European Telecommunications Standards Institute) have also produced standards for smart cards and terminals which are more stringent than that described by the current ISO standards.

Card specification

The specification of a card involves parameters that are common to many existing applications using the ISO ID-1 card. The following list defines the main parameters that should be defined,

- Card dimensions
- Chip location (contact card)
- Card material (e.g PVC,ABS)
- Printing requirements
- Magnetic stripe (optional)
- Signature strip (optional)
- Hologram or photo (optional)
- Embossing (optional)
- Environmental parameters

The characteristics of the smart card are part of the ISO 7816 part 1 (physical) and 2 (contact location) standards. The choice of chip location has been a difficult subject due largely to the use of magnetic stripes. The early French cards put the IC module further off the longitudinal axis of the card than the standard eventually agreed by ISO. This was preferable because of the residual risk of chip damage due to bending. The French Transac tracks were lower on the card which also made this position preferable. The now agreed ISO standards for magnetic stripes resulted in the French chip position and the magnetic stripe being coincident. Hence the now agreed lower location which does of course result in higher bending stress on the chip. The ISO 7816-2 standard does however allow the position of the contacts to be either side of the card. More recently there have been moves to remove this option with the front (opposite to the side containing the magnetic stripe) being the preferred position for the IC connector.

The choice of card material effects the environmental properties of the finished product. PVC was traditionally used in the manufacture of cards and enabled a higher printing resolution. Such cards are laminated as three layers with transparent overlays on the front and back. More recently ABS has been used which allows the card to be produced by an injection moulding process. It is even possible to insert the chip micromodule in one step as part of the moulding process. Temperature stability is clearly important for some applications and ETSI are particularly concerned here, such that their higher temperature requirement will need the use of polycarbonate materials.

Mask ROM Specification

The mask ROM contains the operating system of the smart card. It is largely concerned with the management of data files but it may optionally involve additional features such as cryptographic algorithms (e.g DES). In some ways this is still a relatively immature part of the smart card standards since the early applications used the smart card largely as a data store with some simple security features such as PIN checking. The relevant part of the ISO standard is 7816-4 (commands) and the new work item for part 7 (security). There is a school of thought that envisages substantial changes in this area to account for the needs of multi-application cards where it is essential to provide the necessary security segregation. Finally the developed code is given to the supplier who incorporates this data as part of the chip manufacturing process.

Application Software Specification

This part of the card development process is clearly specific to the particular application. The application code could be designed as part of the mask ROM code but the more modern approach is to design the application software to operate from the PROM non volatile memory. This allows a far more flexible

approach since the application can be loaded into the chip after manufacture. More over by the use of EEPROM it is possible to change this code in a development environment. The manufacture of a chip with the user's ROM code takes on average three months. Application code can be loaded into the PROM memory in minutes with no further reference to the chip manufacturer.

Chip Fabrication

The fabrication of the card involves a number of processes as shown in fig. 2. The first part of the process is to manufacture a substrate which contains the chip. This is often called a COB (Chip On Board) and consists of a glass epoxy connector board on which the chip is bonded to the connectors. There are three technologies available for this process, wire bonding, flip chip processing and tape automated bonding (TAB). In each case the semiconductor wafer manufactured by the semiconductor supplier is diced into individual chips . This may be done by scribing with a diamond tipped point and then pressure rolling the wafers so that it fractures along the scribe lines. More commonly the die are separated from the wafer by the use of a diamond saw. A mylar sheet is stuck to the back of the wafer so that following separation the dice remain attached to the mylar film.

Wire bonding is the most commonly used technique in the manufacture of smart cards. Here a 25uM gold or aluminium wire is bonded to the pads on the chip using ultrasonic or thermo compression bonding. Thermo compression bonding requires the substrate to be maintained at between 150C and 200C. The temperature at the bonding interface can reach 350C. To alleviate these problems thermo sonic bonding is often used which is a combination of the two processes but which operate at lower temperatures.

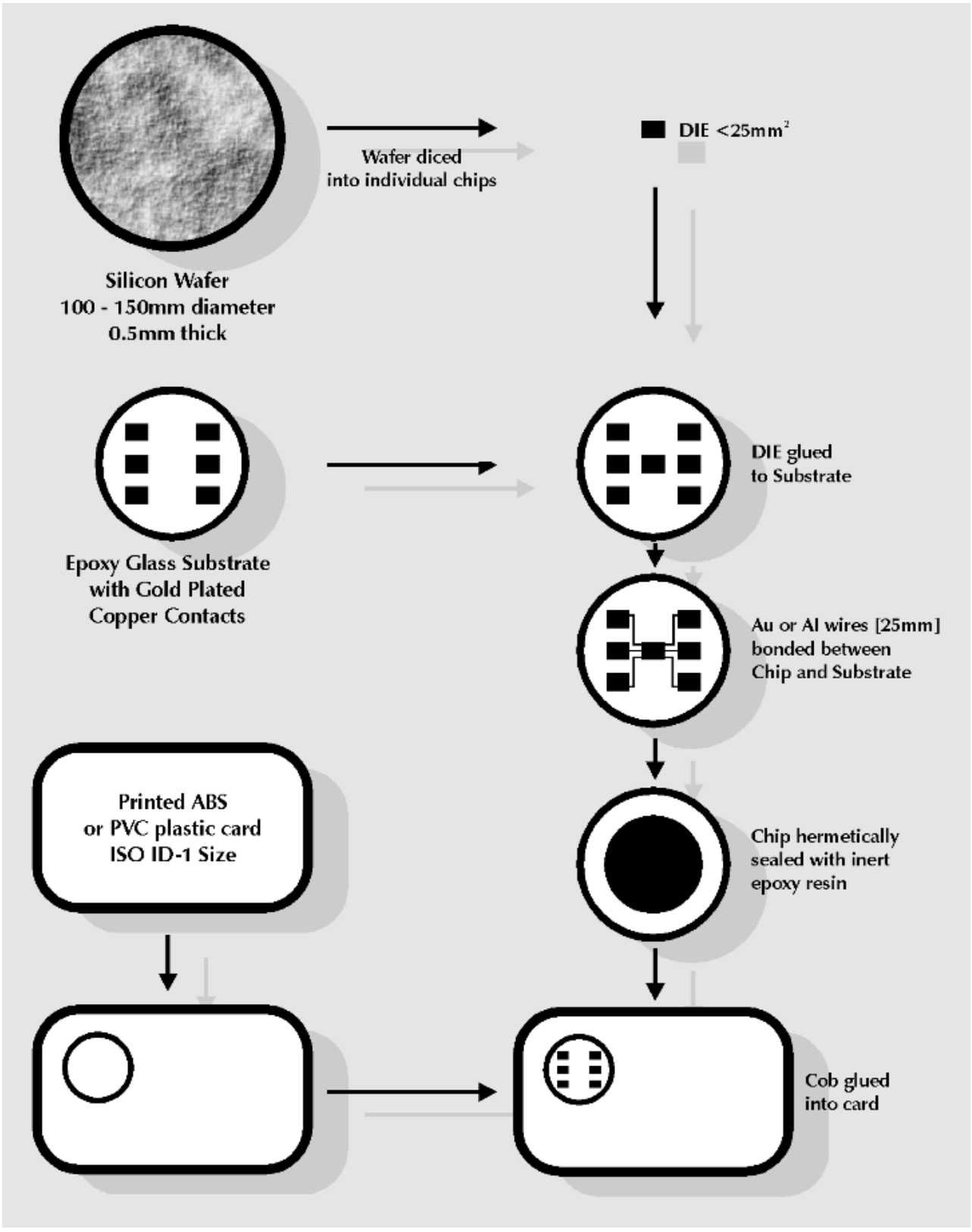


Figure 2: The Manufacturing Process

The die mounting and wire bonding processes involve a large number of operations and are therefore quite expensive. Because in general only 5 or 6 wires are bonded for smart card applications this approach is acceptable. However in the semiconductor industry generally two other techniques are used, the flip chip process and tape automated bonding. In both cases gold bumps are formed on the die. In flip chip processing

the dice are placed face down on the substrate and bonding is effected by solder reflow. With tape automated bonding the dice are attached by thermocompression to copper leads supported on a flexible tape similar to a 35mm film.

The finished substrate is hermetically sealed with an inert material such as epoxy resin. The complete micromodule is then glued into the card which contains the appropriately sized hole.

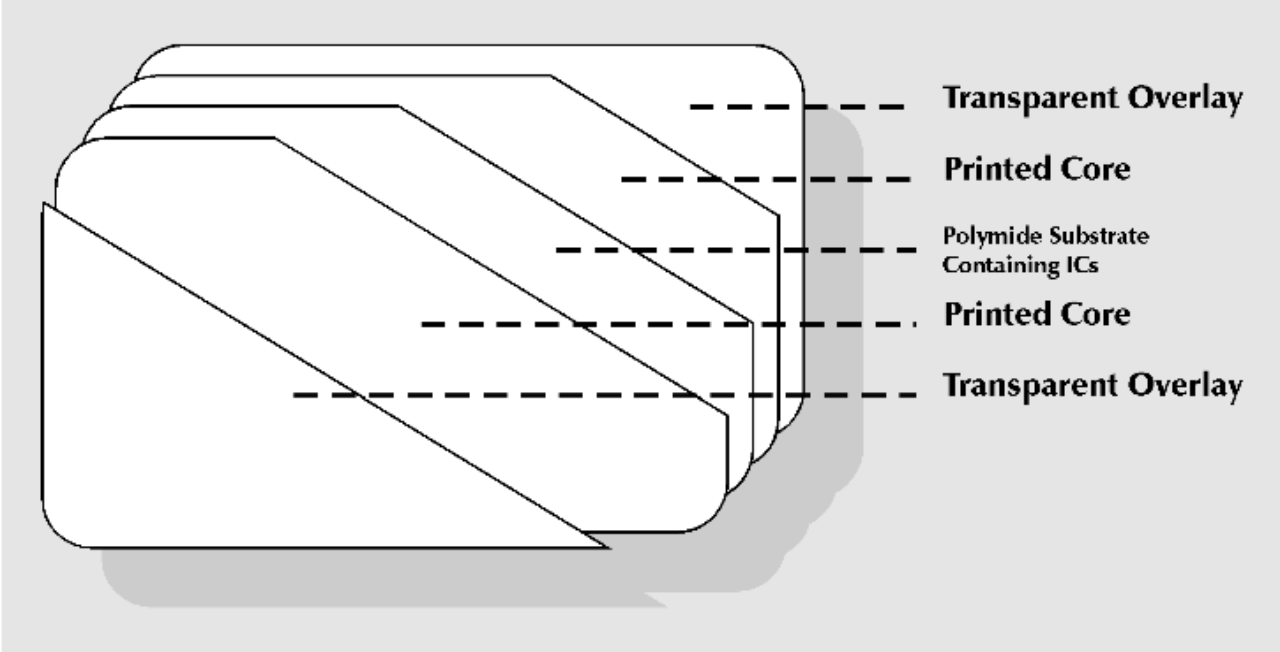


Figure 3: The Contactless Card

The fabrication of a contactless card is somewhat different since it always involves a laminated card as shown in fig. 3. The ICs and their interconnections as well as the aerial circuits are prepared on a flexible polymide substrate.

Application load

Assuming the application is to be placed in the PROM memory of the IC then the next stage in the process is to load the code into the memory.

This is accomplished by using the basic commands contained in the operating system in the mask ROM. These commands allow the reading and writing of the PROM memory.

David Everett

Next week - Part 6 “Let’s get physical”.

Smart Card Tutorial

First Published in July 1995

From there to here Part - 6

Physical characteristics of the Contact Card

Interoperability starts at the bottom. If the card won't fit the slot you have a basic problem and little else matters. This month we will look at the ISO standards surrounding the physical characteristics of the Smart Card (Integrated Circuit Card or ICC in ISO parlance) which will ensure the fundamental basis of interoperability.

Equally important is an understanding of the elements in the design and manufacturing process which effects the reliability of the final card product. There are still many observers in the industry who seem to believe that card failure rates up to 10% are the norm. While this may well have happened in the past, such a concept is a totally unacceptable engineering axiom. Today, under normal wear and tear, well designed ICCs are returning less than 0.1% in field use. But what are the best reliability figures and what are the tests that can be used to monitor and improve performance? This area is currently poorly covered by ISO but we can examine some of the industry techniques and look at some future possibilities.

The physical characteristics of an IC card are defined in ISO 7816 part 1. This standard applies to the ID - 1 identification card specified in ISO 7810 and includes cards which may have embossing or magnetic stripes. While we are all familiar with the use of imprinters to obtain a printed version of the embossed characters on some paper voucher, their viability on an IC card must be questionable. The IC module in a Smart Card is like any other electronic component and is not normally expected to be hit with a hammer at regular intervals. Even the embossing process itself is mechanically stressful and must raise serious doubts over the appropriate migration strategy.

The physical properties of the contact IC card are referenced against earlier card standards and we will look at each of them in turn.

ISO 7810 Identification cards - Physical characteristics (1985)

This standard specifies the physical characteristics of identification cards including card material, construction, characteristics and nominal dimensions for three sizes of cards (ID -1, ID -2 and ID -3). It is the ID -1 card that forms the basis of ISO 7816 -1.

The principal parameters of ISO 7810 are the dimensions of the ID -1 card which are defined to be, 85.6mm x 53.98mm x 0.76mm

ISO 7811 Identification cards - recording techniques (1985)

This standard is in five parts and covers the specification of the magnetic stripe and the card embossing. It is not possible to entirely ignore these standards since devices such as ATMs use the magnetic stripe to open the vandal proof gate on the card slot.

Part 1: Embossing

This part specifies the requirements for embossed characters on identification cards for the transfer of data by imprinters or by visual or machine reading.

Part 2: Magnetic stripe

This part specifies characteristics for a magnetic stripe, the encoding technique and coded character sets which are intended for machine reading.

Part 3: Location of embossed characters on ID -1 cards.

As the title implies, this part of the standard specifies the location of embossed characters on an ID -1 card for which two areas are assigned. Area 1 is for the number identifying both the card issuer and the card holder. Area 2 is provided for the cardholder identification data such as his name and address.

Part 4: Location of magnetic read only tracks - tracks 1 and 2

This standard specifies the location of the magnetic material, the location of the encoded data tracks and the beginning and end of the encoding.

Part 5: Location of read - write magnetic track - track 3

This standard has the same scope as part 4 except that it defines the read - write track 3.

ISO 7812 Identification cards- numbering system and registration procedure for issuer identifers (1987)

This standard relates to the card identification number or PAN (Primary Account Number) which consists of three parts, the issuer identifier number (IIN), the individual account identifier and the check digit.

ISO 7813 Identification cards - Financial transaction cards (1987)

This standard defines the requirements for cards to be used in financial transactions. It specifies the physical characteristics, layout, recording techniques, numbering system and registration procedures. It is defined by reference to ISO 7810, ISO 7811 and ISO 7812.

In particular the standard defines more precisely the physical dimensions of the card as follows,

- Width 85.47mm - 85.72mm
- Height 53.92mm - 54.03mm
- Thickness 0.76mm $\pm$ 0.08mm

The thickness of the card is particularly important for Smart Card readers because of the mechanical construction of the card connector mechanism.

This device often consists of a movable carriage that positions the card under the connector head while applying the necessary wiping and pressure action. Variation in thickness or even slight warping of the card can cause communication failures.

ISO 7816 Design and use of identification cards having integrated circuits with contacts (1987)

This standard in its many parts is probably the most important specification for the lower layers of the IC card. The first five parts in particular are well established and allow total physical and electrical interoperability as well as defining the communication protocol between the IC card and the CAD (Card Acceptor Device).

Part 4 defines the command structure for communicating with a Smart Card while part 5 describes the identification and registration of applications.

Part 1: Physical characteristics

The physical dimensions of the IC card are defined as that specified in ISO 7813. It should be noted that the thickness dimension does not include any allowance for embossing. More particularly the slot for a card may include an extra indentation for the embossed area of the card. In effect it acts as a polarisation key and may be used to aid the correct insertion orientation of the card. This is an additional characteristic to the magnetic field sensor which operates off the magnetic stripe and is used to open a mechanical gate on ATM devices as mentioned previously.

The part 1 standard also defines additional characteristics that should be met in the manufacture of an IC card. These characteristics fall into the following categories,

- Ultra violet light
- X - rays
- Surface profile of contacts
- Mechanical strength (of cards and contacts)
- Electrical resistance (of contacts)
- Electromagnetic interference (between magnetic stripe and integrated circuit)
- Electromagnetic field
- Static electricity
- Heat dissipation

It has to be said that this part of the standard could be improved and ISO 10373 (Identification Cards - Test methods) goes somewhat further in describing the test methodologies in some detail. The three most widely used tests applied by fabricators are specified in the annex to the standard,

- A1 Bending properties
- A2 Torsion properties
- A3 Static electricity

While this is certainly one way of comparing cards fabricated by different companies, whether it bears any relationship to the use of IC cards in the field seems debatable.

The bending properties are tested by deflecting the card on each axis as shown in fig1. With a periodicity of 30 bendings per minute the card is deflected to 2 cm at its centre from the long axis and 1 cm from the short axis. The recommended test requires the card to withstand 250 bendings in each of the four possible orientations (i.e. 1000 bendings in total).

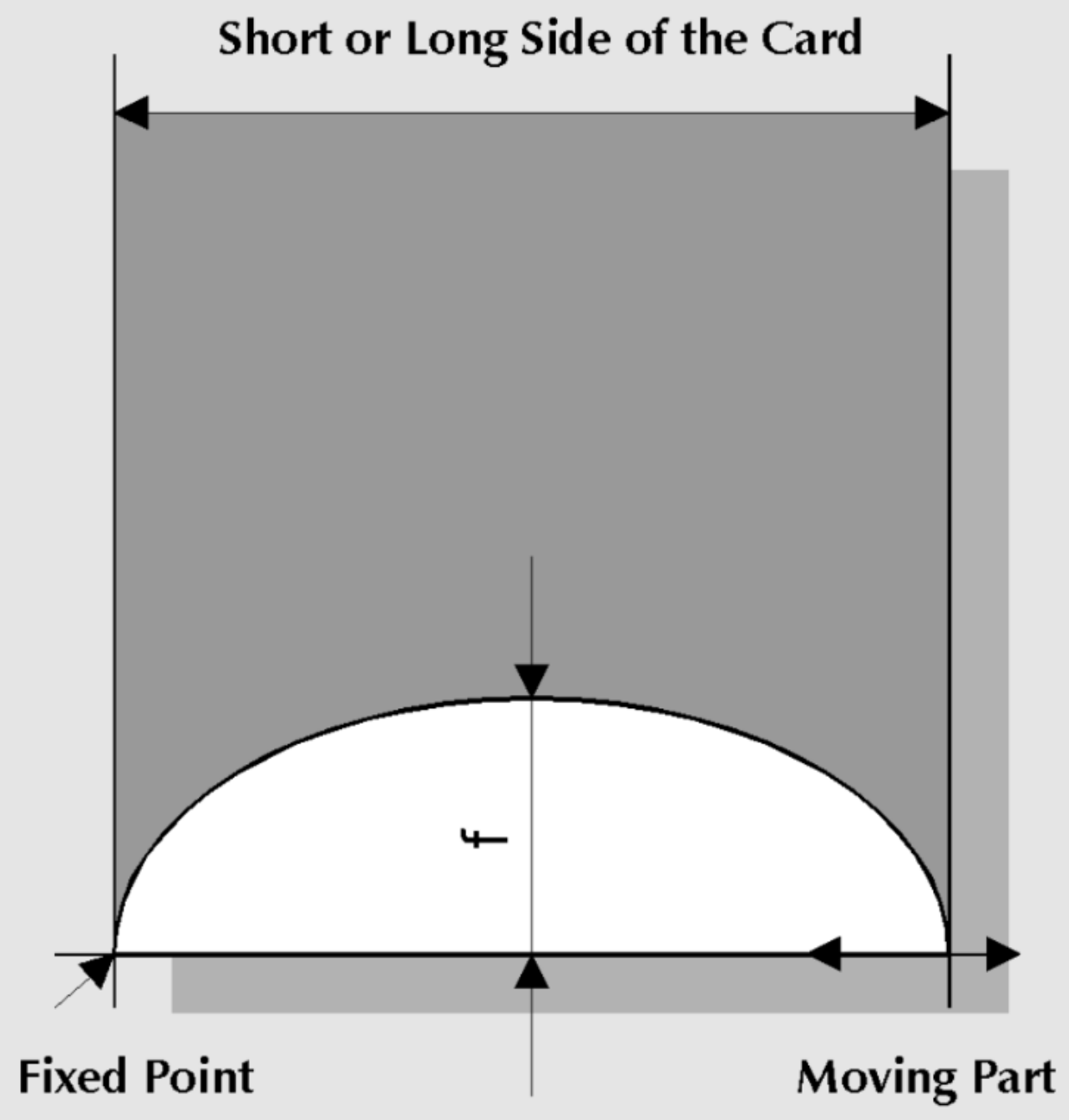


Figure 1

The torsion properties of the card are tested by displacing the card $\pm$ 15^{o} about the long axis at a periodicity of 30 torsions per minute (fig 2). The standard requires the card to withstand 1000 torsions without chip failure or visible cracking of the card.

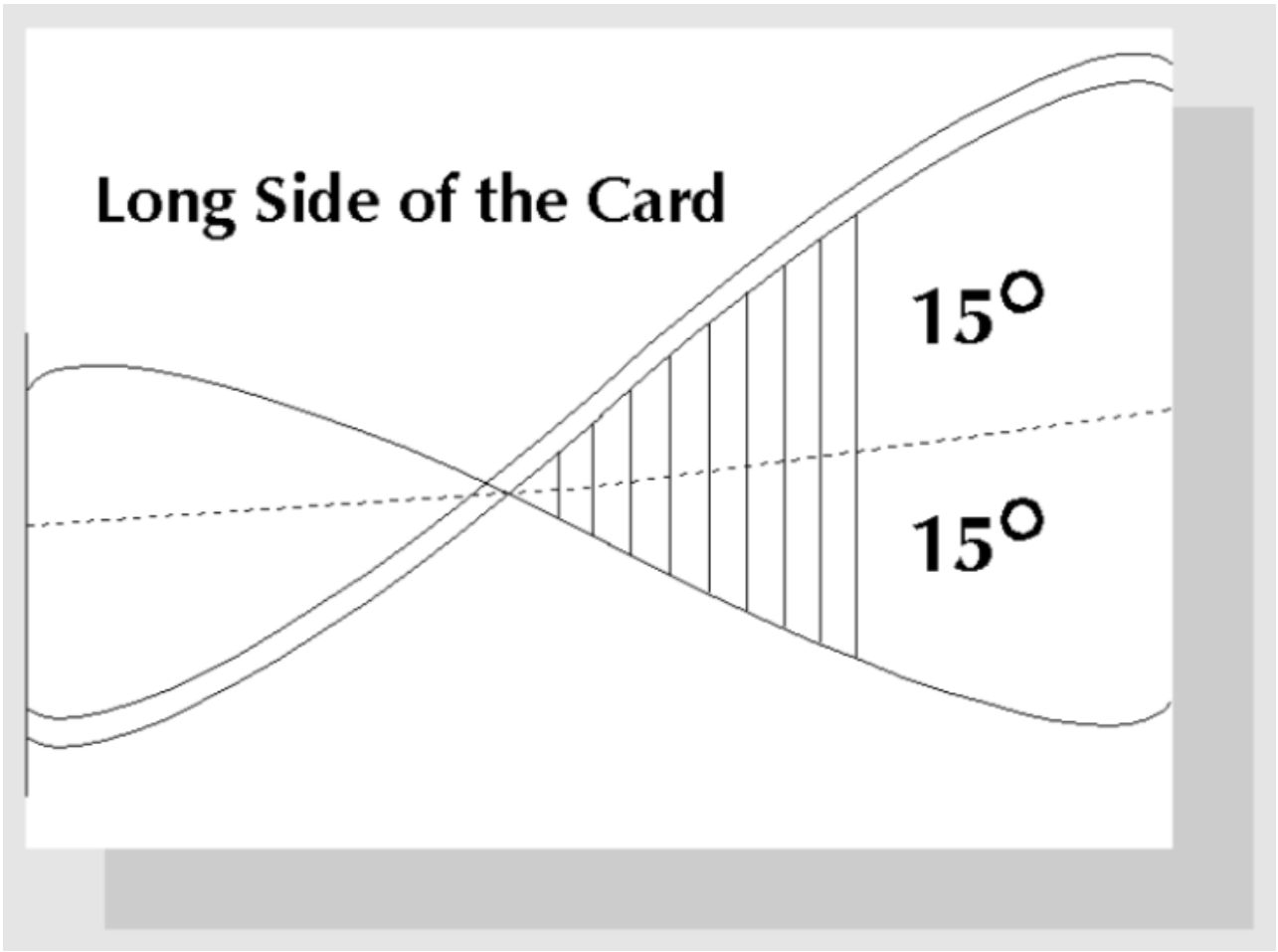


Figure 2

The resistance of the card to static electricity is defined by a test set up as shown in fig 3. The test voltage is defined to be 1.5KVolts. The specification requires this voltage to be discharged across each of the contacts in both normal and reverse polarity. The IC should still be operational at the end of the test.

One of the issues surrounding the use of the IC card relates to the temperature range for operational use. ISO 7810 defines that the ID-1 card should be structurally reliable and usable between -35oC and +50oC. The draft CEN standard on requirements for IC cards and terminals for telecommunications use, part 2 - application independent card requirements (EN 726-2) defines more stringent requirements for operational use as -25oC to +65oC with occasional peaks up to +70oC. In addition the draft identifies multi-application cards for portable battery operated equipment to be used between -25oC and +70oC with occasional peaks of up to +85oC. The word occasional is defined to mean not more than 4 hours each time and not over 100 times during the life of the card.

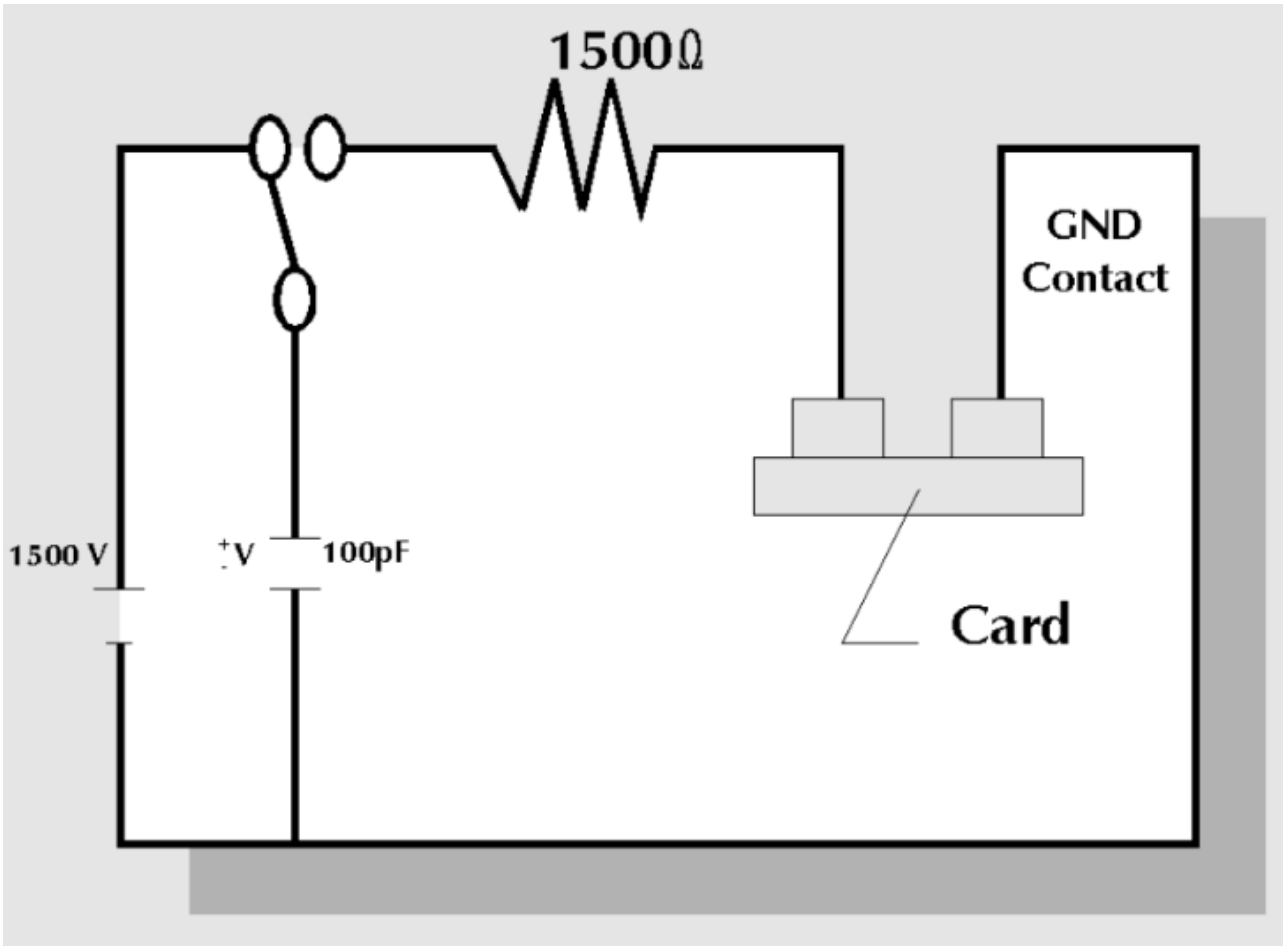


Figure 3

ISO 7816 Part 2 - Contact Locations and Minimum Size

This part of the standard has taken a lot of effort in order to reach agreement. Early applications of Smart Cards emanated in France where the Transac magnetic stripes were more central on the card than that eventually defined by ISO 7811. Unfortunately the French chip position overlaps the ISO magnetic stripe definition. As a result it was eventually agreed that after a transitional period (to the end of 1990) the position for the IC connector would be as shown in fig 4. This position is much closer to the longitudinal axis of the card. We might like to conjecture on which is the better position for the chip in terms of mechanical stress but perhaps we should just settle for agreement.

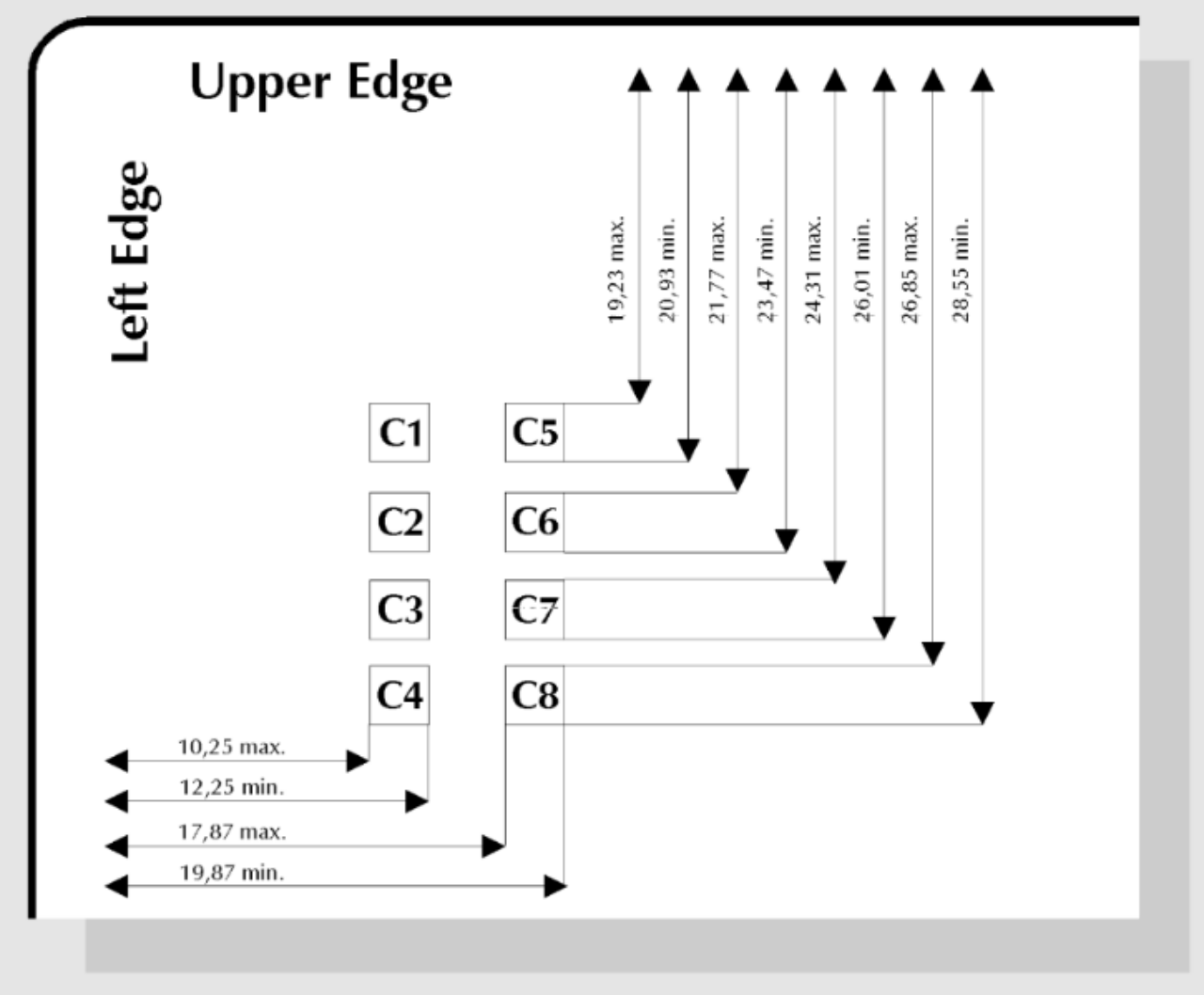


Figure 4

Further problems arose in deciding on which face of the card the connector should be located. In order to avoid further delay in publishing the standard, two options were allowed to include both the front and back of the card. This anomaly has been a source of irritation and it is now widely agreed that the IC connector should be on the front of the card. For this purpose the back is defined to be the side with the magnetic stripe. The embossing is defined to be on the front of the card and therefore on the same side as the IC connector. The relative location of these components (when present) is shown in fig 5.

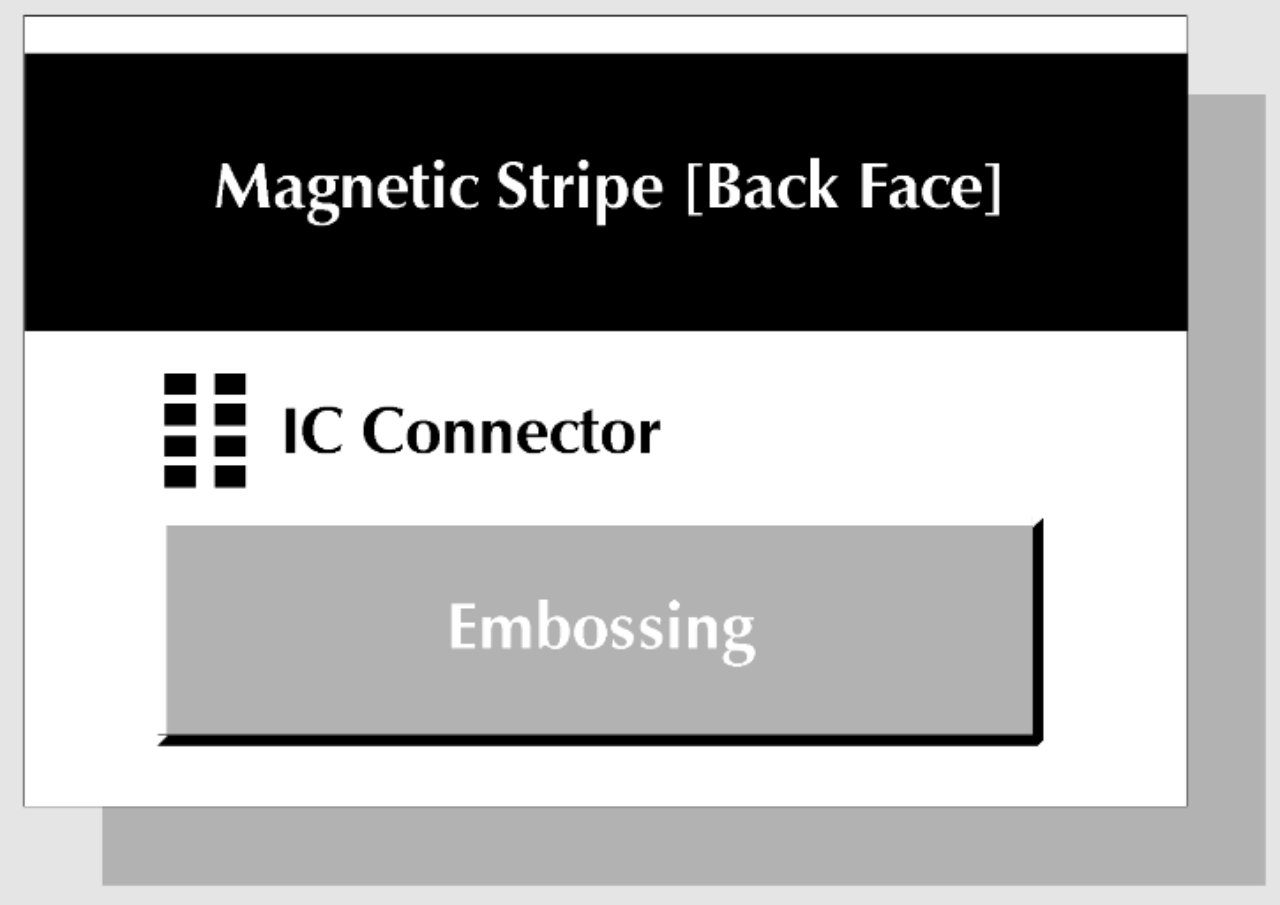


Figure 5

Reliability in the field

At the end of the day what we really want apart from interoperability is field reliability. It is generally accepted by experts in the subject that the tests referred to so far are somewhat inadequate. We are not suggesting that these tests are inappropriate only that we need a new view on the subject.

In practice with cards produced by the experienced fabricator we are unlikely to see plastic card failure due to bending or torsion in the way described in the tests. Yet the various postal authorities have designed automatic letter handling machines where the count path follows a bend that used to demolish the chip module with unfailing regularity. Many manufacturers have developed their own bending test where the card is forced around cylinders of various diameters to emulate the postal problem. Arguably this test is far more significant than those defined by ISO.

Failure due to electro static damage (ESD) these days is insignificant due to the improvement in chip and packaging design. Most ICCs are capable of a much higher performance than that defined in the ISO standard.

So where are all these card failures? In fact, they are largely due to excessive stress on the micromodule or the package that contains the chip. This module is sometimes referred to as the COB which stands for Chip On Board and relates to one form of packaging where the chip is stuck on a miniature printed circuit board and bond wires are used to connect the chip to the connecter plate. This technique is still widely used.

We propose that the final reliability of the card is largely a function of the physical characteristics of the chip packaging. Accordingly the primary test methods for Smart Cards should be aimed at the micromodule. One of the major failure modes is due to pressure perpendicular to the plane of the module. Various test

techniques to examine the properties of the module have been developed by the various card manufacturers, but they are all aimed at measuring resistance to a force either applied continuously or by an impact impulse. Just to bring in the practical relevance of such forces, just have a look at your own card wallet. Does it have one of those pop fasteners? Why does it always seem to line up with the centre of the chip?

David Everett (next month - part 7)

Smart Card Tutorial

First Published in August 1995

From there to here Part - 7

Electronic Signals and Transmission Protocols.

There is currently much debate about the interoperability of the various Smart Card schemes being developed around the world. The picture of multiple terminals on the retailers counter is still fresh in the minds of those people involved in the development of EFTPOS (Electronic Funds Transfer at the Point Of Sale) terminals. The implication of various standards and specifications is often misunderstood so in this part of the tutorial we will try to resolve the key anomalies.

In general we can consider the scenario where the application running in one host computer (the chip in the Smart Card) and an application running in another host computer (the terminal) wish to exchange data. Clearly both the Smart Card and the terminal may contain several different applications. There is no reason to expect that these applications are in any way interoperable. This may be electronic purse applications, medical applications or even retailer loyalty applications. Unless the same application exists in both host computers, the meaningful exchange of application data is not valid. These applications are normally implemented by software modules running on both host computers. Clearly the hosts may be technically different as may the software modules but they are functionally compatible.

The International Standards Organisation ISO has developed an OSI (Open Systems Interconnection) model which defines a number of layers (7) by which data can be exchanged between applications running on systems that are "open" to each other by means of mutually agreeable standards. We can apply the same approach to Smart Card systems but can avoid most of the complexity of the OSI model which is not relevant to a much simpler implementation.

Smart Card applications are usually invoked as a simple command/response architecture using a direct link between the two hosts. As such our greatest interest here is to ensure interoperability that will allow data exchange to use a common standard for the lowest two layers of the OSI model, the physical layer and the data link layer.

It is primarily these two layers which are the subject of ISO 7816-3 which is the standard for electronic signals and transmission protocols.

The physical layer covers four concepts,

- Mechanical
- Electrical
- Functional
- Procedural

We have already discussed the mechanical specification so our attention will be directed to the electrical, functional, and procedural aspects of the ISO 7816-3 standard. A well known physical layer standard is RS232-C. This is not actually an ISO standard but came from the Electronic Industries Association of Washington DC (EIA). The International counterpart is V24 which comes from the International Telegraph and Telephone Consultative Committee (CCITT). The Smart Card standard ISO7816-3 does not conform to either RS232-C or V24.

Before we start on the detail of the ISO 7816-3 standard we need to inform readers that this part of the standard is currently under review. One of the briefs is to incorporate the two additional ammendments produced since the original standard was agreed concerning the T=1 communications protocol and protocol type selection (PTS). The significant changes are in the voltage and current supply to the Smart Card where

the voltage supply range is to be increased from 5V only to allow 3V to 5V operation. The other significant change surrounds the current supply available from the terminal to the Smart Card. The existing standard specifics 200 mA capacity but this is likely to be decreased to 50mA. This in part represents reality (no common Smart Card takes more than 50mA) and the need to move towards more viable battery operation in terminals.

The electronic properties and transmission characteristics of the IC card are fundamental to interoperability. The principal subjects to be considered are as follows,

- Electrical characteristics
- Character transmission
- Answer to reset (ATR)
- T=0 transmission protocol
- T=1 transmission protocol
- Protocol type selection (PTS)

We will consider each of these topics in turn.

IC Card Electrical Characteristics

We have previously discussed the position and definition of the IC connector and have identified 8 contacts of which 6 are currently defined,

- V_{CC} Power supply
- GND Ground or reference voltage
- CLK Clock
- V_{PP} Programming voltage
- RST Reset signal
- I/O Serial Input/Output

Power supply (V_{CC})

The power supply to the IC is defined to be between 4.75 volts and 5.25 volts with a maximum current consumption of 200mA. Both of these parameters have problems. Newer chip fabrication technologies are moving sub micron, 0.8um is already commercially available and 0.5um is not that far away. These chips may operate with a supply voltage of 3 volts which results in lower current consumption. Most card acceptor devices (CAD) operate at 5 volts as specified in the ISO standard. Whilst a 3 volt IC may be designed to operate between 3 volts and 5 volts, running a 5 volt IC at 3 volts may be a non starter.

A current consumption of 200mA is far too high for modern electronic equipment particularly when the equipment is portable and driven by a battery power supply. Most IC cards have a power consumption of between 10mA and 20mA (at 3.58MHz). ETSI in the development of their standards have adopted a far more rigorous specification of 20mA maximum for normal use and a 10mA maximum for use in portable equipment. They further defined the concept of sleep mode (not covered by ISO 7816-3) where the IC chip can reside in a latent mode preserving volatile memory contents with a maximum power consumption of 200uA.

Clock signal

Although the integrated circuit could contain its own clock circuit for driving the internal logic, in practice most IC chips are supplied with an external clock by the interface device. It should be noted that the speed of the serial communications on the I/O line is effectively defined by the frequency of this clock. The ISO standard aligns with the use of two widely used external clock frequencies, 3.579545 MHz and 4.9152 MHz. The former frequency is the more widely used (being based on the NTSC colour sub carrier frequency) and results in a clock divider of 372 in order to produce a 9600 bit per second (not exact but within tolerance)

serial communication speed. The latter frequency has a simple divisor of 512 in order to achieve a 9600 bit per second communication speed. The standard defines the situation after reset whilst allowing the frequency to be selectively changed by means of protocol type selection.

Programming voltage V_{PP}

This signal is designed to provide the high voltage required to enable writing to the non volatile memory. The more popular IC'c use EEPROM memory where the high voltage is generated by a charge pump on chip. However the EPROM memory type needs the high voltage (usually 12.5V or 21V) to be externally provided on the IC connector. There have been problems in the past with terminals supplying the wrong programming voltage with somewhat drastic effects. Because of this and the significant advantages of having a rewriteable memory the EEPROM memory is by far the most popular for IC card applications, hence the role of VPP is rapidly diminishing.

The Reset Signal

The reset signal is asserted by the interface device and is used to start up the program contained in the IC ROM. The ISO standard defines three reset modes, internal reset, active low reset and synchronous high active reset. Most microprocessor ICs operate using the active low reset mode were the IC transfers control to the entry address for the program when the reset signal returns to the high voltage level. The synchronous mode of operation is more commonly met with the memory card ICs as used for telephone applications.

The sequence of operations for activating and deactivating the IC is defined in order to minimise the likelihood of damage to the IC. In particular the inadvertent corruption of the non-volatile memory (EPROM or EEPROM) must be avoided. The activation sequence for the interface device is defined as follows,

- Take RST low
- Apply V_{CC}
- Put I/O in receive mode
- Put V_{PP} in idle mode
- Apply clock
- Take RST high (active low reset)

The IC deactivation sequence for the interface device is as follows,

- Take RST low
- Take clock low
- Deactivate V_{PP}
- Put I/O in the low state
- Deactivate V_{CC}

Serial Input/Output (I/O)

The ISO standard defines a single line for the interchange of data between the IC and the interface device. This means that the line must change direction depending on whether the IC is transmitting or receiving. In practice this cannot be instantaneous and the expression 'line turnaround time' is commonly encountered in the modem world. The transmission protocol must take account of this need to turn the line around.

Character Transmission.

The transmission characteristics operated by most microprocessor IC cards are based on an asynchronous half duplex mode of operation. In the T=0 communication protocol this involves the transmission of bytes whilst the T=1 protocol defines a block mode of operation. As we have already observed the serial communication is operated by the use of a single chip connector, where the direction of data transmission has to change depending on whether the IC card or interface is transmitting data. This is referred to as half

duplex communication whereas two I/O signal connectors would be required for full duplex operation where transmission can take place in both directions concurrently.

The asynchronous type of transmission is similar to that used by the serial RS232C connector met on the personal computer. Although the PC operates in full duplex mode. The transmission of a single character (defined as 8 bits) requires an overhead of several bits as follows,

- Start bit (used for character frame synchronisation)
- Parity bit (for error detection)
- Guardtime (separation between characters)

The format of a character frame is shown in fig.1 The receiver examines the I/O line looking for the transition from the mark or high state to the space or low state. The sampling of the line is required to be such that the receiver monitors the state of the line in the centre of each bit period with a precision of + 20%. The parity bit is defined to achieve even parity which means that the number of 1's in the 8 data bits and the parity bit together results in an even number.

The guard time is defined to be equal to two bit periods (although for block mode it can be changed to a 1 bit period). This is similar to having two stop bits on a UART (Universal Asynchronous Receiver Transmitter) as used in the PC.

A more common definition of the asynchronous serial transmission at reset would be 9600 bits/second, 8 data bits, even parity, 2 stop bits with half duplex mode of operation. The half duplex refers only to data transmissions in one direction at a time which a PC is perfectly capable of managing with its UART. The RS232C interface however defines two separate wires for data transmission and reception which would need hardware modification in order to interface with the single wire IC card directly.

There is a further problem with the asynchronous character transmission that makes life difficult for a PC to act as the interface device. The 7816-3 standard defines an error detection and recovery operation (mandatory for T=0) that cannot be managed by the normal PC UART. When the receiver detects a parity error on reception it takes the I/O line to the space or low state in the middle of the first stop bit guard time. The transmitter is mandated to sample the I/O line at the start of the second stop bit guard time period. When the error condition is sensed then the transmitter should retransmit the erroneously received character. Clearly the transmitter cannot be outputting stop bits but must let the line go high during the guard time in order to sense the line state. Given the close coupling normally achieved between an IC card and the interface device one has to question whether this level of error control has sufficient benefits to outweigh the disadvantages. Error control at a higher level in the OSI model is preferable in this situation and although this could be handled at the application level the T=1 communication protocol applies error control at the frame level.

David Everett (next month - part 8)

Electronic Signals and Transmission Protocols - continued.

Smart Card Tutorial

First Published in September 1995

From There to Here - part 8

Electronic Signals and Transmission Protocols - continued

The asynchronous character frame (fig. 1) is at the heart of the communications system for Smart Cards. It looks remarkably simple but we have probably all experienced the problems of using the RS 232C communications bit and the undeniably subtle problems that can take hours to fix.

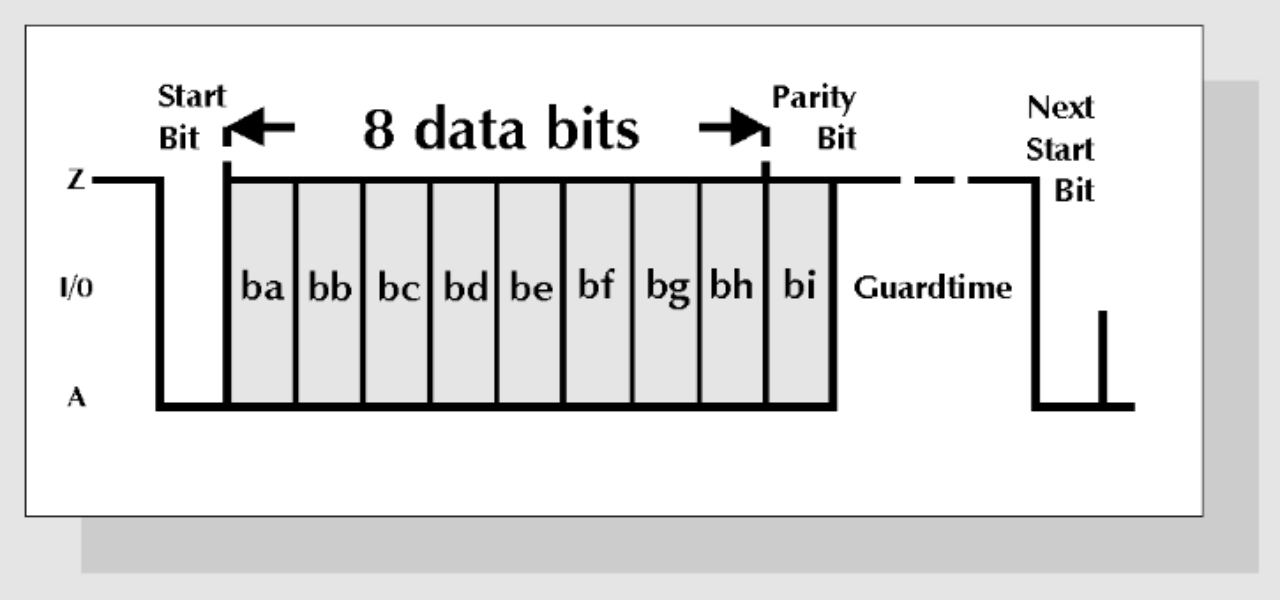


Figure 1: Asynchronous Frame

This month we will just tidy up the error detection and correction used with the byte asynchronous (T=0) communication protocol and then go and look at the answer to reset. Although these protocols are well established, very few cards and terminals strictly obey the standard, but then very few devices correctly obey all the aspects of RS232C. In practice, life is based on de-facto standards which are all an agreed interpretation of a significant part of the International Standard for interoperability (under normal conditions) to be achieved.

The T=0 error detection mechanism described in ISO 7816-3 is mandatory for the card and optional for the terminal. It's operation can be followed by reference to figure 1.

When the receiver detects a parity error on reception it takes the I/O line to the space or low state in the middle of the first stop bit guard time. The transmitter is mandated to sample the I/O line at the start of the second stop bit guard time period. When the error condition is sensed then the transmitter should retransmit the erroneously received character. Clearly the transmitter cannot be outputting stop bits but must let the line go high during the guard time in order to sense the line state. Given the close coupling normally achieved between an IC card and the interface device one has to question whether this level of error control has sufficient benefits to outweigh the disadvantages. Error control at a higher level in the OSI model is preferable in this situation and although this could be handled at the application level the T=1 communication protocol applies error control at the frame level.

Answer to reset

After the reset signal is applied by the interface device the IC card responds with an answer to reset. For the active low reset mode the IC should respond between 400 and 40,000 clock cycles after the rising edge of the reset signal. The answer to reset is at most 33 characters (including the initial character) and consists of 5 fields,

- The initial character (TS)
- The format character (TO)
- The interface characters (TA_i,TB_i,TC_i,TD_i,)
- The historical characters (T1, T2.TK)
- The check character (TCK)

Each of these fields is sent in order as shown in fig.2. The initial character TS is really a bit synchronisation pattern which may be sent in order to determine the data transmission rate (auto baud rate sensing) and also to determine the sense of the logic. The format of the TS character is shown in fig.3. This shows the two possibilities of the direct and inverse convention. In the inverse convention where the logic level 1 is the space or low state the most significant bit is transmitted first. With the direct convention where the logic level 1 is the mark or high state then the least significant bit is transmitted first. This means that the selection of the appropriate logic sense will result in the initial character being interpreted as `3F' for the inverse convention and `3B' for the direct convention in hexadecimal coding.

The format character TO provides information necessary to interpret the remaining answer to reset characters. The most significant 4 bits use a bit map to indicate the presence or otherwise of TA_i, TB_i, TC_i and TD_i. For example if the most significant bit (b8) is set then TD1 is present in the interface characters field. Similarly the presence of TC_i is indicated by the state of the `b7' bit and so on.

The least significant 4 bits of the TO formal character give the number (binary encoded) of bytes in the historical field. The use of 4 bits restricts the maximum size of the historical character field to 15 bytes.

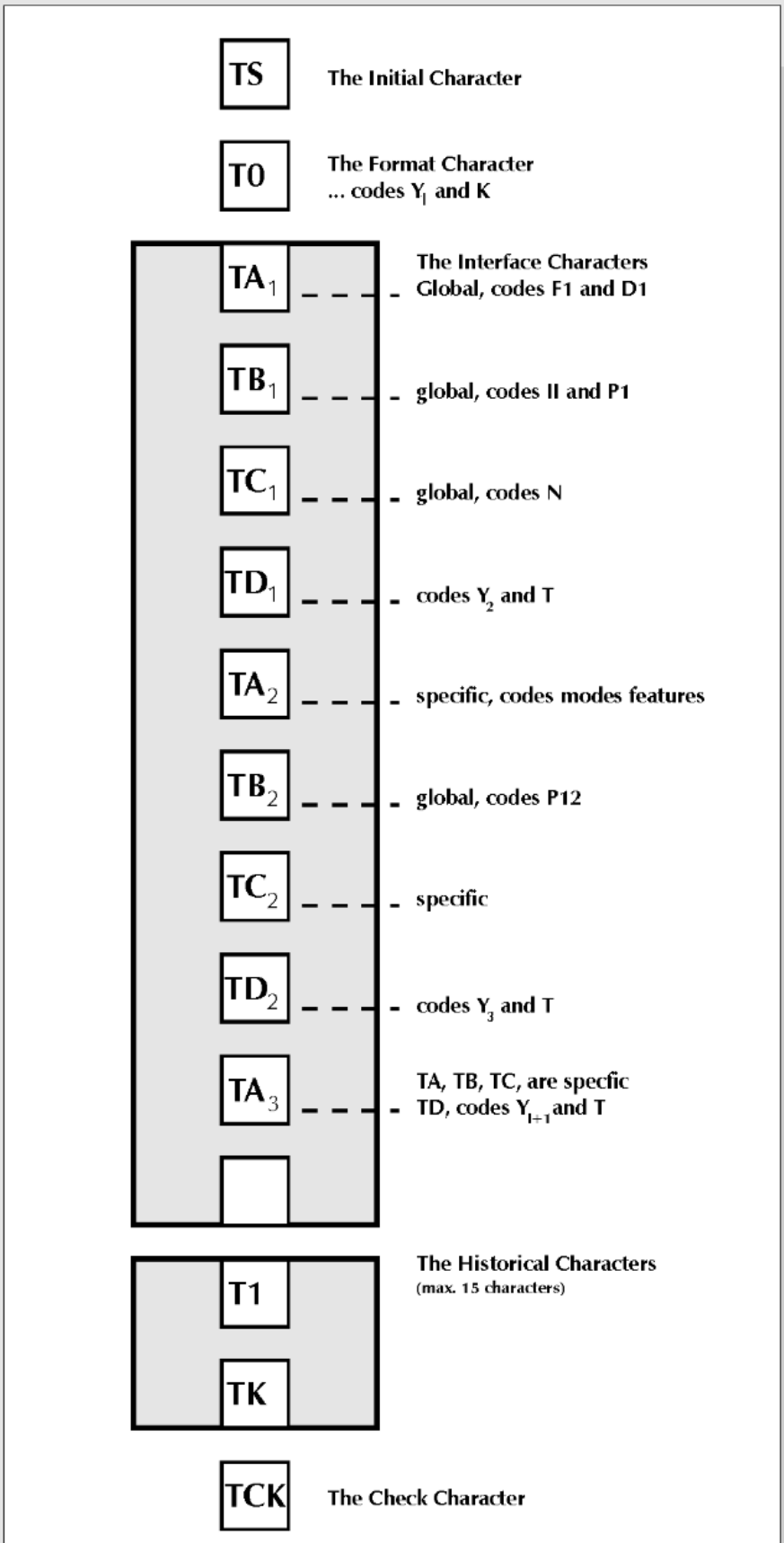

TS
The Initial Character
T0
The Format Character
... codes Y1 and K
TA1
The Interface Characters
Global, codes F1 and D1
TB1
global, codes II and P1
TC1
global, codes N
TD1
codes Y2 and T
TA2
specific, codes modes features
TB2
global, codes P12
TC2
specific
TD2
codes Y3 and T
TA3
TA, TB, TC, are specfic
TD, codes Yi+1 and T
T1
The Historical Characters
(max. 15 characters)
TK
TCK
The Check Character

Figure 2: General Configuration of the Answer-to-Reset

The interface characters (TA_i, TBi, TC_i, TD_i,) are the complex part of the answer to reset. They carry information relating to the available communication protocols as well as the programming voltage and current parameters for the EPROM. There is currently a proposed revision to the ISO 7816-3 to remove ambiguities and to ensure an effective method of operation for changing the protocol type and the protocol parameters. Much of the complexity is brought about by the desire to achieve backward compatibility with commercial implementations of the T=O communication protocol. At the current time there are commercial applications running either the T=O or T=1 communication protocol whilst multi-protocol operation is somewhat scarce.

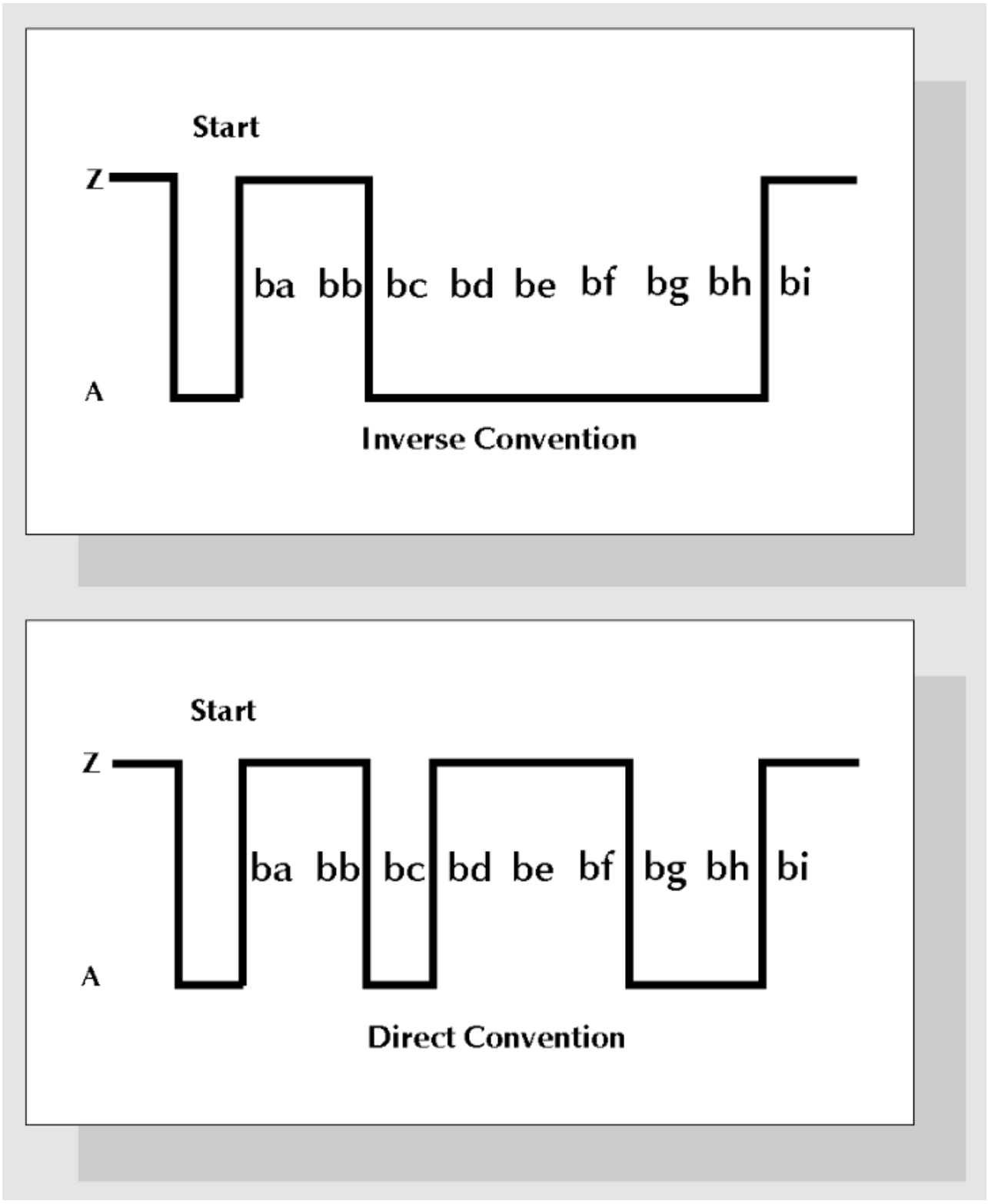


Figure 3: Initial Character TS

The proposed revisions to the standard may alter this situation. We will discuss the interface bytes and protocol type selection against these proposed revisions but readers are warned that these recommendations are only provisional.

The interface bytes (which are optional) are defined in fig.4. The T_0 and TD_i characters contain bit maps which indicate the presence or otherwise of the following TA_i, TB_i, TC_i, and TD_i bytes.

The TA_i, TB_i, TC_i, and TB_2 characters are referred to as the global interface bytes and are fundamental to the operation of the card.

TA_i defines the basic characters of the serial transmission, FI is the clock rate conversion factor and DI is the bit rate adjustment factor. The binary encoded fields are compared against tables supplied in the standard to achieve actual values for F and D as defined below,

$$Work\ etu = \frac{1}{D} \times \frac{F}{f} \sec$$

An elementary time unit (etu) is the nominal bit duration used in the character frame. Thus as described previously one character frame is equal to 12 etu (1 start etu, 8 data etu, 1 parity etu, 2 guard time etu).

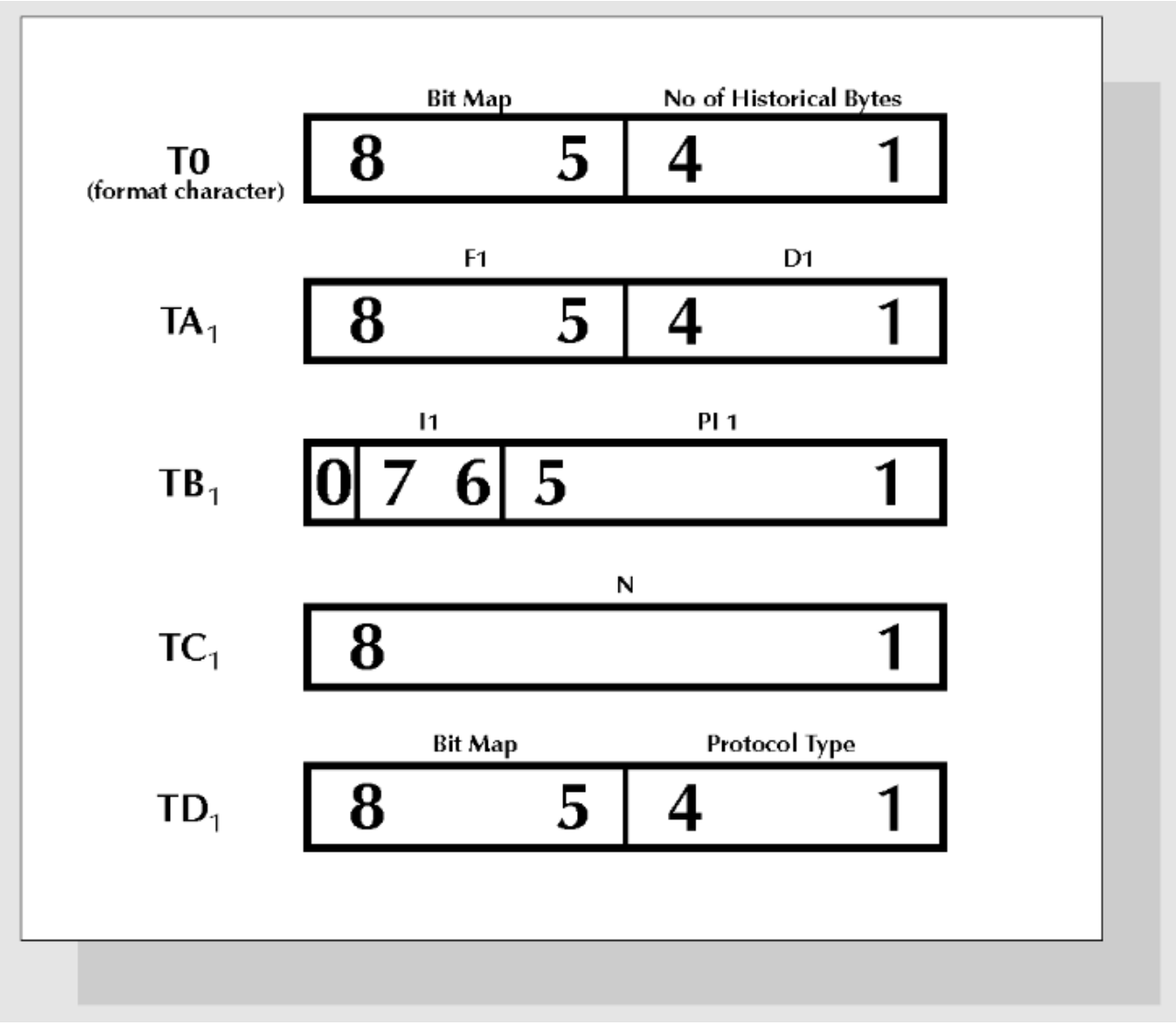


Figure 4: The Interface Bytes

The default values for F1 and D1 are 1 which is defined in the tables to give a value for F of 372 and D of 1. Hence the work and initial etu are the same. At these default values the frequency of the clock should be in the range 1MHz - 5MHz.

TB_i is used to define the EPROM programming voltage and current. The value of II and PI1 are used against tables to obtain the value of I mA and P volts. It should be noted that TB_2 is used to define the programming voltage with higher granularity (8 bits instead of 5).

TC_i provides the value of N which defines the extra guard time to be used between successive characters. N can be in the range 0 - 254 etu. When N is equal to 255 this indicates that the minimum guard time (2 etu for T = 0 and 1 etu for T = 1) should be used. As noted previously the T = 0 communications protocol requires the extra guard time to enable the parity error detection and signalling to be implemented.

TD_i indicates the protocol type TD_i as between 0 and 15,

- T = 0 Asynchronous half duplex byte transmission
- T = 1 Asynchronous half duplex block transmission
- T = 2/3 Reserved for full duplex operation
- T = 4 Reserved for enhanced half duplex byte transmission
- T = 5..13 Reserved for further use (RFU)
- T = 14 Non ISO protocols
- T = 15 Reserved for future extension

It should be noted that Japan uses T = 14 for a National block asynchronous protocol.

The TD_i byte also contains a bit map that indicates the presence or otherwise of TA_2, TB_2, TC_2 and TD_2.

The proposed revision defines a new use for the TA_2 interface byte which has a special role in the selection of communication protocols and parameters. We will discuss this further in the communications section.

The Historical Characters

The historical characters may be used to convey information relating to the life cycle of the card. There are clearly other possibilities and the use of these characters is still subject to agreement. This subject is being considered further as part of the emerging part 4 of the ISO 7816 standard.

The Check Character (TCK)

The check character should not be sent when only the T = 0 protocol is indicated in the answer to reset. In all other cases TCK is sent as the last character of the ATR. The check character is calculated such that the Exclusive OR of all the bytes from T0 to TCK inclusive is equal to zero.

David Everett (next month part 9) - Transmission protocols continued.

Smart Card Tutorial

First Published in October 1995

From There to Here - Part 9

The T=O Communication's Protocol

Communication protocols are an emotive subject and Smart Card communication is little different. The ISO 7816-3 standard allows for the card to hold multiple protocols and provides a method of switching between them. In practice this is a debatable concept since at the very least it loads the processor and memory of the Smart Card to the advantage of the terminal.

The ISO standard currently defines two communication protocols,

- T=0 asynchronous half duplex character transmission
- T=1 asynchronous half duplex block transmission

The T=0 protocol is relatively simple and has the lowest overhead on the Smart Card chip. The T=1 protocol is more sophisticated and includes much improved error handling. You can argue the right and wrongs until the cows come home but the difference is simply in the overhead on the chip. The T=1 protocol is better from a pure communication point of view but requires significantly more memory in the chip, both program memory and RAM working space. The designer needs to assess the reliability of the communication link and whether error handling at the application level is more appropriate.

The T = 0 protocol is still the predominant protocol and was the original protocol specified in ISO 7816 - 3. In 1992 ISO standardised the T = 1 protocol as amendment 1 to ISO 7816 - 3. Clearly the IC card and the interface device must operate with a common protocol. The method by which they achieve a common optimum configuration has been the subject of much discussion over the last few years. This principle is intended to be achieved by the use of protocol type selection (PTS). This is effectively a special command sent from the interface device to the ICC after the answer to reset. In order to maintain backward compatibility with existing commercial systems that may only be capable of handling the T=0 communication protocol some changes are necessary to the original ISO 7816-3 standard. A new concept is proposed which identifies the principle of two modes of operation,

- Negotiable mode
- Specific mode

An ICC that operates in a negotiable mode may have its communication protocol changed by the use of the PTS command. An ICC that operates in the specific mode cannot accept a PTS command but may be put into the negotiable mode by a further assertion of the reset command.

Although the ICC indicates to the interface device (by means of TA_2) its capability to change to the negotiable mode, an existing device in the market place may however be unaware of these changes and therefore will not be prepared to reset the card.

The operation of these mode changes are shown in fig.1. It should be noted that a multi protocol card which by definition offers the negotiable mode of operation should give priority to the T=0 communication protocol. In other words if the T=0 protocol is available it should be the default protocol offered in the answer to reset.

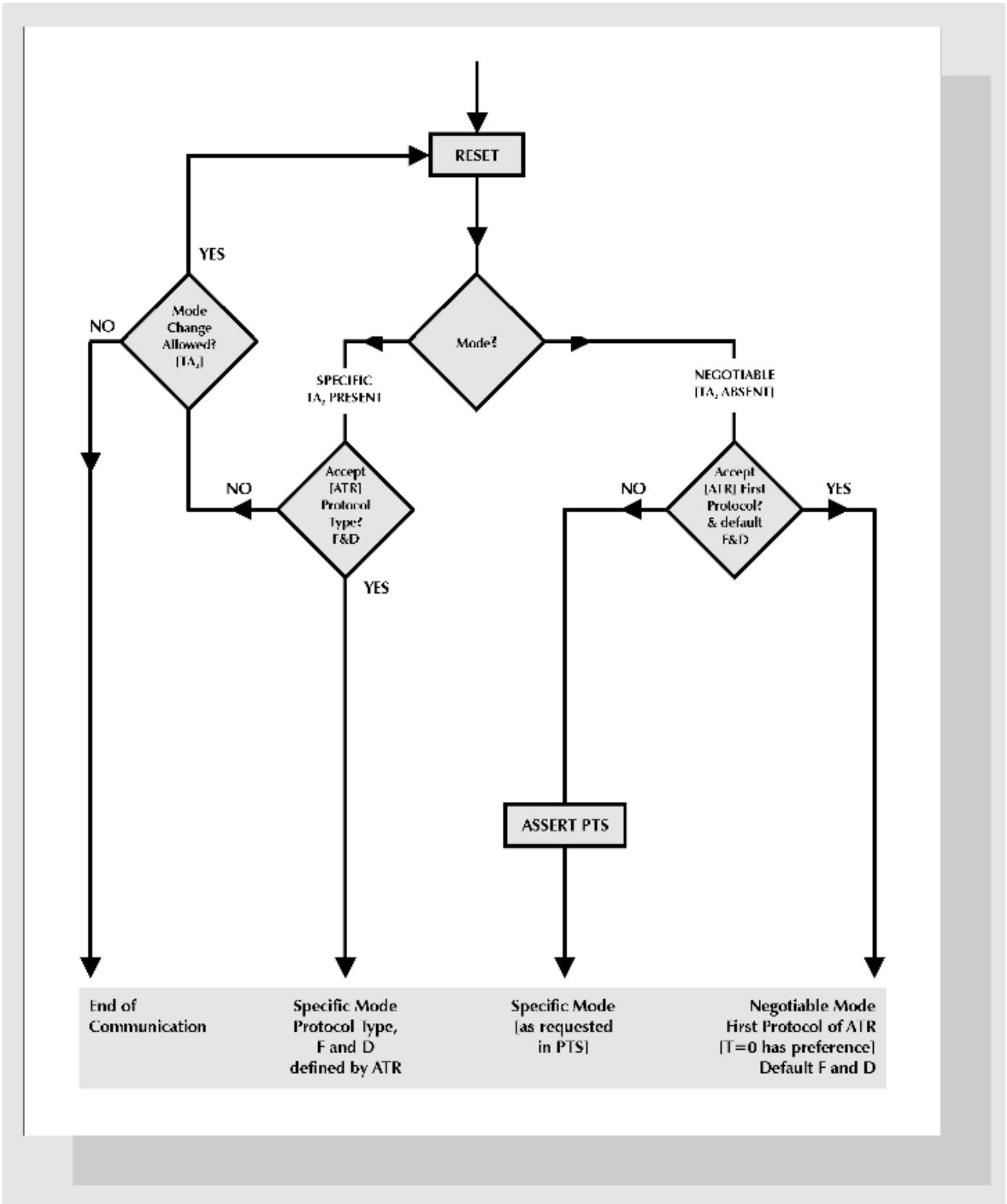


Figure 1: Modes of Operation

The TA_2 interface byte which is part of the answer to reset data (discussed in part 4) gives the necessary information to allow the appropriate choice of protocol. The coding of this byte when present is shown in fig.2. In fact the presence or otherwise of this byte is used to determine the mode of operation of the card as follows,

- TA_2 present in ATR - Specific mode

- TA_2 absent in ATR - Negotiable mode

It can be seen that bit 8 in the TA_2 byte is used to tell the interface device whether the card can change to the negotiable mode.

Protocol Type selection (PTS)

Protocol type selection is used by the interface device to change the communications protocol and/or the default values of FI and DI. The PTS command must be issued immediately after the answer to reset and only applies when the IC card is in the negotiable mode.

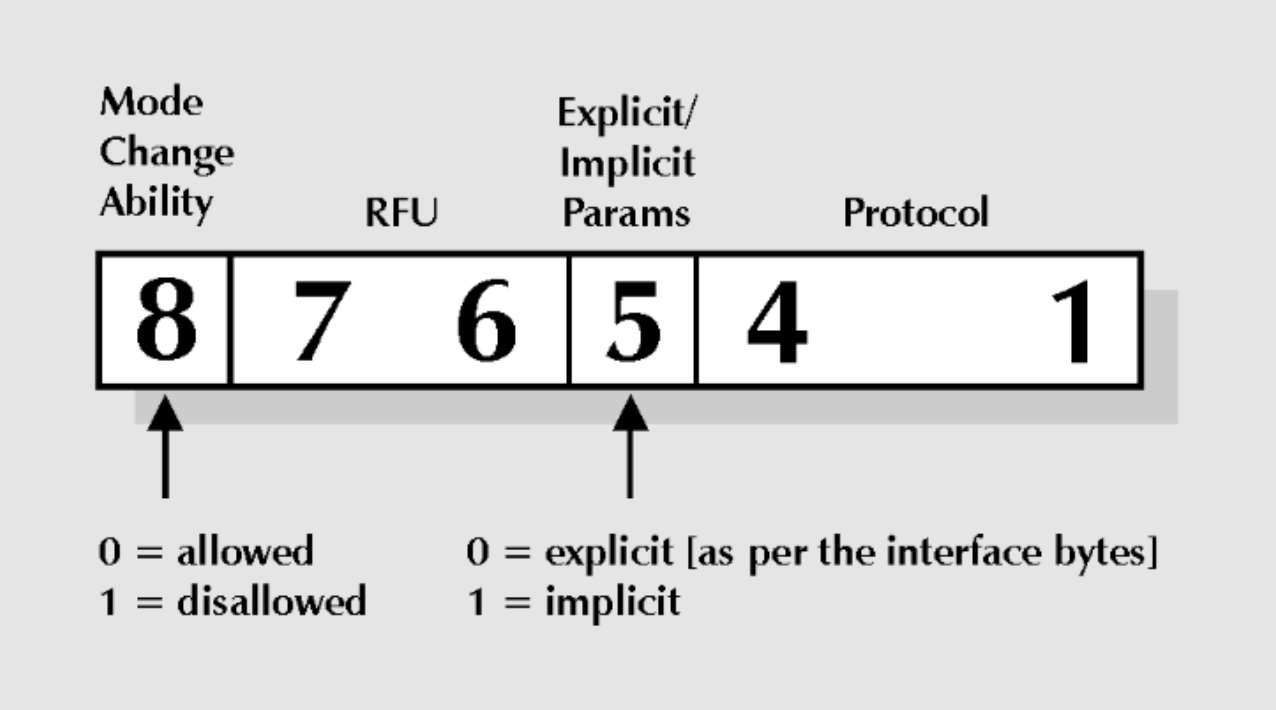


Figure 2: The TA2 Interface Bytes

The interface device may choose to operate by using the first indicated protocol after the answer to reset and by using the default values of F and D. This results in an implicit selection of the protocol and the communication parameters. Should the interface device wish to effect any change to this situation then it must issue the PTS command.

The PTS request consists of an initial character PTSS (coded FF_{hex}), followed by a format character PTSO, and three optional characters PTS1, PTS2, PTS3 and PCK the check character. This is shown in fig.17. The response from the ICC follows the same format as the request.

The PTS0 format character is encoded as shown in fig.3. The bit map is used to indicate the presence or otherwise of PTS1, PTS2 and PTS3. These are encoded by bits 5, 6 and 7 respectively where a logic `1' level indicates the presence of the character. The protocol type is indicated by bits 1, 2, 3 and 4 which are binary encoded for T=0 to T=15.

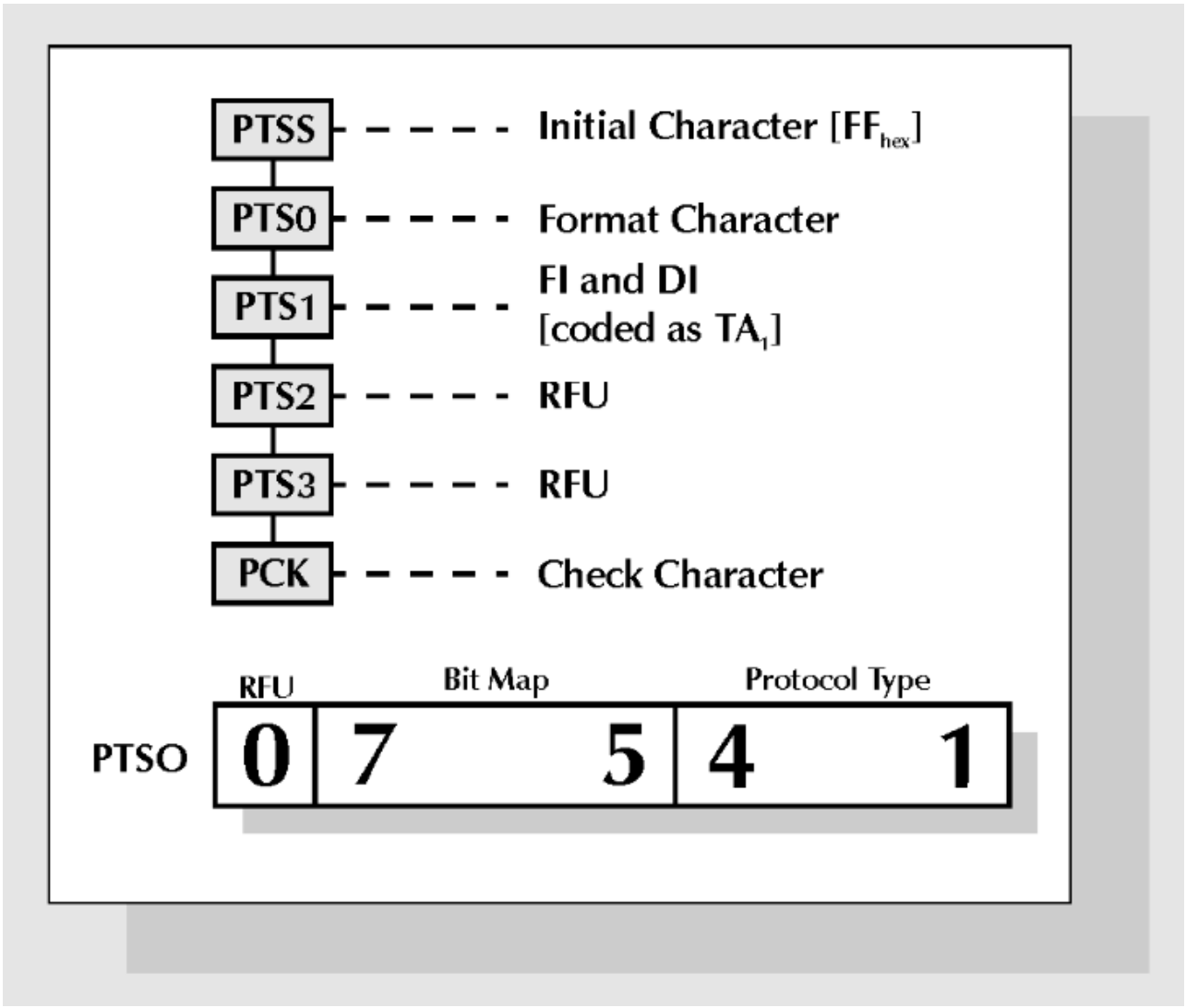


Figure 3: PTS Request and Response

The PTS1 character when present is used to define the values for FI as coded for TA1. These parameters are used for defining the work etu (elementary time unit).

The check character PCK is computed such that the exclusive OR (XOR) of all the characters from PTSS to PCK inclusive is equal to zero.

When the ICC implements the PTS request message correctly it replies by echoing the same request as the response message. If bit 5 of the PTS1 response character is set to zero then the default values of F and D will be used.

The T=0 communication protocol

The interface device always initiates the command for the T=0 protocol. Interaction between the interface device and the ICC results in successive commands and responses. For this protocol, data can only flow in one direction for the command response pair. In other words, either the command message contains data for the ICC or the command request data from the ICC which is then included in the response. The direction of data flow is implicit on the definition of the command and hence both the interface device and the ICC need to have the necessary a-priori knowledge. When it is required to transfer data in both directions for a particular command then a get response command may be used after the primary command to recover the response data.

The command message consists of a 5 character header which the interface device sends to the ICC. The ICC then replies with a procedure byte after which either data is sent to the ICC, or from the ICC, depending on the particular command. This procedure byte is to allow the interface device to control the V_{pp} EPROM programming voltage. In the case of EEPROM memory this procedure byte is effectively redundant. The message flow for the T=0 protocol is shown in fig.4. The command header consists of the following 5 bytes,

- CLA - the instruction class (FF is reserved for PTS)
- INS - the instruction code (eg read memory)
- P1 - instruction code qualifier (eg memory address)
- P2 - additional INS code qualifier
- P3 - the length of the data block

When P3 is equal to zero the data from the card will be 256 bytes. When data is to be transferred into the card then a zero data transfer is implied.

The normal condition for the ACK procedure byte is for this byte to echo the instruction byte (INS). Other options allow the interface devices to control the $_{Vpp}$ programming voltage as required. The card may optionally send a NULL procedure byte (60hex) which allows further time for the processing of the command. In this situation the IFD should await a further procedure byte. The ISO standard also allows the card to send the first status byte (SW1) as the procedure byte.

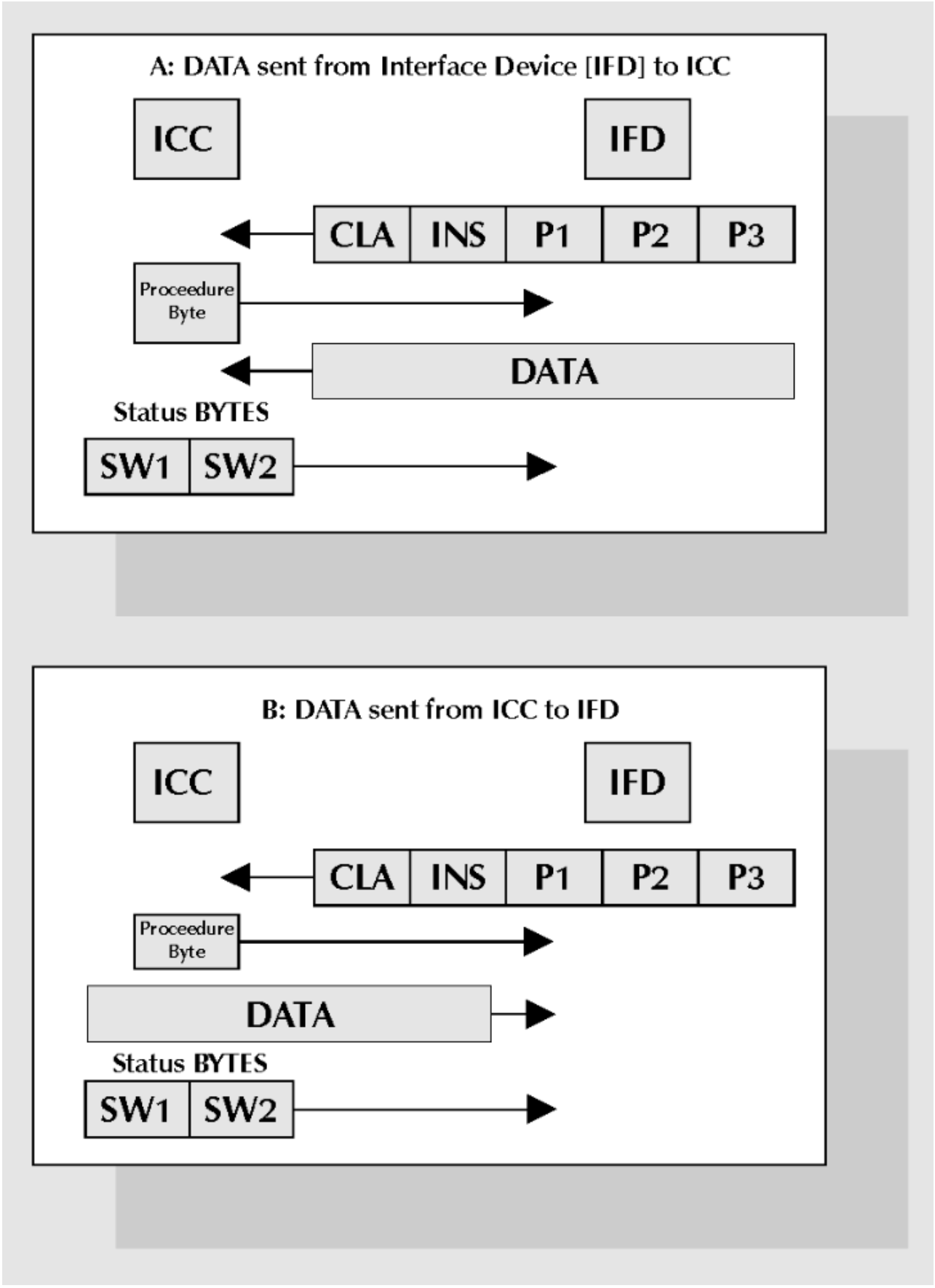


Figure 4: The T=0 Message Protocol

There are two status bytes SW1 and SW2. These bytes are sent from the ICC to the interface device on completion of the command to indicate the current card status. The normal response is,

- SW1, SW2 = 90_{hex}, 00_{hex}

When SW1 = 6X or 9X various error conditions are reported by the card. ISO 7816-3 defines 5 such error conditions,

- SW1=6E - Card does not support instruction class
  = 6D - Invalid INS code
- SW1= 6B - Incorrect reference
  = 67 - incorrect length
  = 6F - no particular diagnosis

The T=0 protocol also includes an error detection and correction mechanism. This was described in part 4 and relies on the receiver detecting a parity error upon which it takes the I/O line to the low logic level within the first etu guard time (10.5 $\pm$ 0.2 etu) for a minimum of 1 etu and a maximum of 2 etu. The transmitter looks for this condition and retransmits the corrupt character.

David Everett

(next month - part 10) - The T=1 communication protocol

Smart Card Tutorial

First Published in November 1995

From there to here - part 10

The T = 1 Communications Protocol

Carrying on from last month, we can now proceed to look at the more comprehensive block protocol.

The T = 1 communication is an asynchronous half duplex block transmission protocol. In terms of the OSI model this protocol operates at layer 2, the data link layer. The physical layer (layer 1) operates in the same way as for the T = 0 protocol except for the error detection and correction. In essence this protocol puts an envelope around a block of characters which allows,

- flow control
- block chaining
- error correction.

The choice of communication protocol for the ICC is still a hot topic and one has to consider what advantages can be offered by the block protocol and then to examine the price that must be paid.

The most obvious advantage of the T = 1 protocol is the ability to manage data flow in both directions. In our discussion of the T = 0 protocol it was shown that for a particular command that the data is either sent to or received from the ICC. This limitation was really due to the use of a single byte for defining the length of the data related to the command.

The T = 1 protocol also removes the T = 0 restriction of the master slave relationship where the interface device (IFD) always initiates a command to which the ICC responds. For this block protocol a command may be initiated by either the IFD or the ICC albeit within the restrictions of the protocol.

A further advantage of the T = 1 protocol is the ability to chain the blocks of data such that an arbitrarily large block of data may be transferred as the result of a single command by the transmission of the appropriate number of frames chained in sequence.

The block protocol also has a more sophisticated error management system. This allows the use of a block error detection code (EDC) and the ability to re-transmit blocks that are subject to some error condition. By comparison the T = 0 protocol has a primitive character error detection and correction scheme as described previously in the tutorial (part 4).

Clearly there is a price to be paid for this higher layer protocol. Apart from the more complex software in both the ICC and the IFD the protocol is more demanding on the RAM memory of the ICC which needs to maintain the last sent block in case retransmission is required. In general the T = 1 protocol offers advantages where the application is managing large blocks of data, particularly when it is required to pass data in both directions as part of a particular command. The efficiency of the protocol is only really apparent for larger data transmissions since the underlying physical layer is still operating in character mode as for the T = 0 protocol. The reduction of the character frame to 11 etu (elementary time units) compared with the 12 etu demanded by T = 0 has to be balanced against the administrative overhead of the frame structure which has both a prologue and epilogue.

There can be no doubt that the error control is significantly improved over the T = 0 protocol but at the lower speed of 9600 bit/second operated by many ICC's over very short transmission paths the probability of communication errors is much reduced. However it is clear that there is a move towards the use of the T = 1 protocol and it seems highly likely that this will become the predominant protocol of the future. We should

not however dismiss the use of the T = 0 protocol which in some situations may well offer a more optimum technical solution. The T = 1 protocol is specified in the ISO standard ISO 7816 - 3 / AMD.1

The block frame

The frame consists of three fields,

- prologue field
- information field (optional)
- epilogue field

as shown below.

Prologue Field			Information Field	Epilogue Field
Node Address NAD	Protocol Control Byte PCB	Length LEN	Optional INF	Error Detection LRC or CRC EDC
1 Byte	1 Byte	1 Byte	0-254 Bytes	½ Bytes

The prologue field consists of three bytes,

- NAD the node address
- PCB protocol control byte
- LEN the data length

The NAD byte uses bits 3 -1 to identify the source address and bits 7 - 5 to identify the destination address. The bits 4 and 8 are used for V_{pp} control which will not be discussed further here. The node address byte allows the use of multiple logical channels where required otherwise both addresses should be set to zero.

The PCB byte allows the identification of three types of block frame,

- An information block (I - block)
- A receive ready block (R - block)
- A supervisory block (S - block)

The information block is the frame which is used to transmit application commands and data between the ICC and the IFD. The receive - ready block is used as an acknowledgement when the protocol is sending data as a sequence of chained blocks. The supervising block is used to establish control parameters and to effect a resynchronisation or abort status as the result of some error condition. The information block also acts as an acknowledgement byte in the non chaining mode.

The LEN byte indicates the number of bytes (if any) in the information field of the frame. Its allowed range of values are from 00 - FE_{hex}. This allows a maximum information field of 254 bytes.

The information field is used to convey the application commands and data which we will discuss in the next part of the tutorial.

The epilogue field contains the block error detection code which may be either an LRC (longitudinal redundancy check) or a CRC (cyclic redundancy check). The LRC is 1 byte whilst the CRC occupies 2 bytes. This option is defined by the specific interface characters.

Specific Interface Characters.

In a previous part of the tutorial (part 4) we discussed the specific interface characters given by the answer to reset (ATR). The T = 1 protocol uses three of these characters to establish the necessary options before communication can take place. These bytes are assigned as follows (where I > 2),

- TA_i = IFSC (default = 32)

- TB_i
  (bit 4 - 1) = CWI (default = 13)
  (bit 8 - 5) = BWI (default = 4)

- TC_i
  (bit 1 = 1) = CRC option
  (bit 1 = 0) = LRC option (default)

The IFSC is the information field size for the card. There is also an IFSD which is the information field size for the interface device. This has a default value of 32 bytes and can only be changed by means of an S - block request from the IFD to the ICC.

Waiting Times

The T = 1 protocol uses two waiting time parameters to help flow control,

- Character Waiting Time (CWT)
- Block Waiting Time (BWT)

The character waiting time is the maximum time between successive characters in a block whilst the block waiting time is the maximum time between the leading edge of the last character in a block sent be the IFD and the leading character of the next block sent by the card.

The character waiting time may be used to detect an error in the length of a block whilst the block waiting time may be used to detect an unresponsive card. There is also a block guard time (BGT) which is defined as the minimum time between the leading edge of the last character of one block and the leading edge of the first character in the new block to be sent in the alternative direction. The CWT and BWT are calculated from the values of CWI and BWI coded as shown previously in the specific interface bytes by means of the following equations,

- $CWT = (2^{CWI} + 11)$ etu
- $BWT = (2^{BWI} \times 960 \times 372 / f)$ Sec + 11 etu

Where f is the clock frequency.

The minimum value for the BWT is 100 mS + 11 etu when the card operates with the default frequency of 3.58 MHZ. The block guard time has a value of 22 etu such that the delay between the start of the last character of a received block and the start of a transmitted block is greater than BGT but less than BWT. Accordingly the minimum inter block time is 11 etu which is equal to one character time.

David Everett (next month part 11 - Transmission protocols continued)

Smart Card Tutorial

First Published in December 1995

From there to here - part 11

The T = 1 Communications Protocol - continued

Protocol control byte

This month we will complete the T=1 communications protocol by looking at the range of possibilities for the Protocol Control Byte (PCB) and then, more particularly, we will look at T=1 in action.

The protocol control byte identifies the different types of block and carries some control information including a single bit sequence number (N) and a block chaining bit (M). Other bits are used to identify transmission errors. The PCB is coded as in the table below:

The I blocks can occur as independent blocks or as part of a chain. The "More" bit is set to indicate that further blocks are to follow. The sequence number of the sender alternates between '0' and '1' starting with '0'.

The R blocks are used to acknowledge the successful or otherwise receipt of transmitted blocks. The sequence number N carries the value of the next expected value of N where the transmitter alternates the value as mentioned above.

While blocks transmitted as part of a chain must be acknowledged by an R block the receipt of a successful stand alone I block may be acknowledged by an I block response. The two correspondents manage the sequence numbers of their I blocks independently alternating between '0' and '1'. The R block has three possible states as shown in the table.

The S blocks are used to invoke four control states as shown in the table. The resynch request is used by the IFD (only) to force a reset of the block transmission parameters to their initial values. A chain may be aborted by either the IFD or ICC perhaps due to some physical error such as memory corruption. The ICC may send an IFS request to invoke a change in the IFSC it can support. Similarly the IFD may send an IFS request to indicate a new IFSD it can support. The S block control also allows the ICC to request an extension to the block waiting time (BWT) that may be necessary to execute a command received in an I block. The INF field in this block contains a single byte integer value which is to be calculated as a multiple of the BWT value. In all cases the receiver of an S block should send the appropriate response block.

The T = 1 Protocol in operation

Using the notation of the ISO 7816 standard we can show the basic operation of the protocol. A more complete definition can be obtained from the standard.

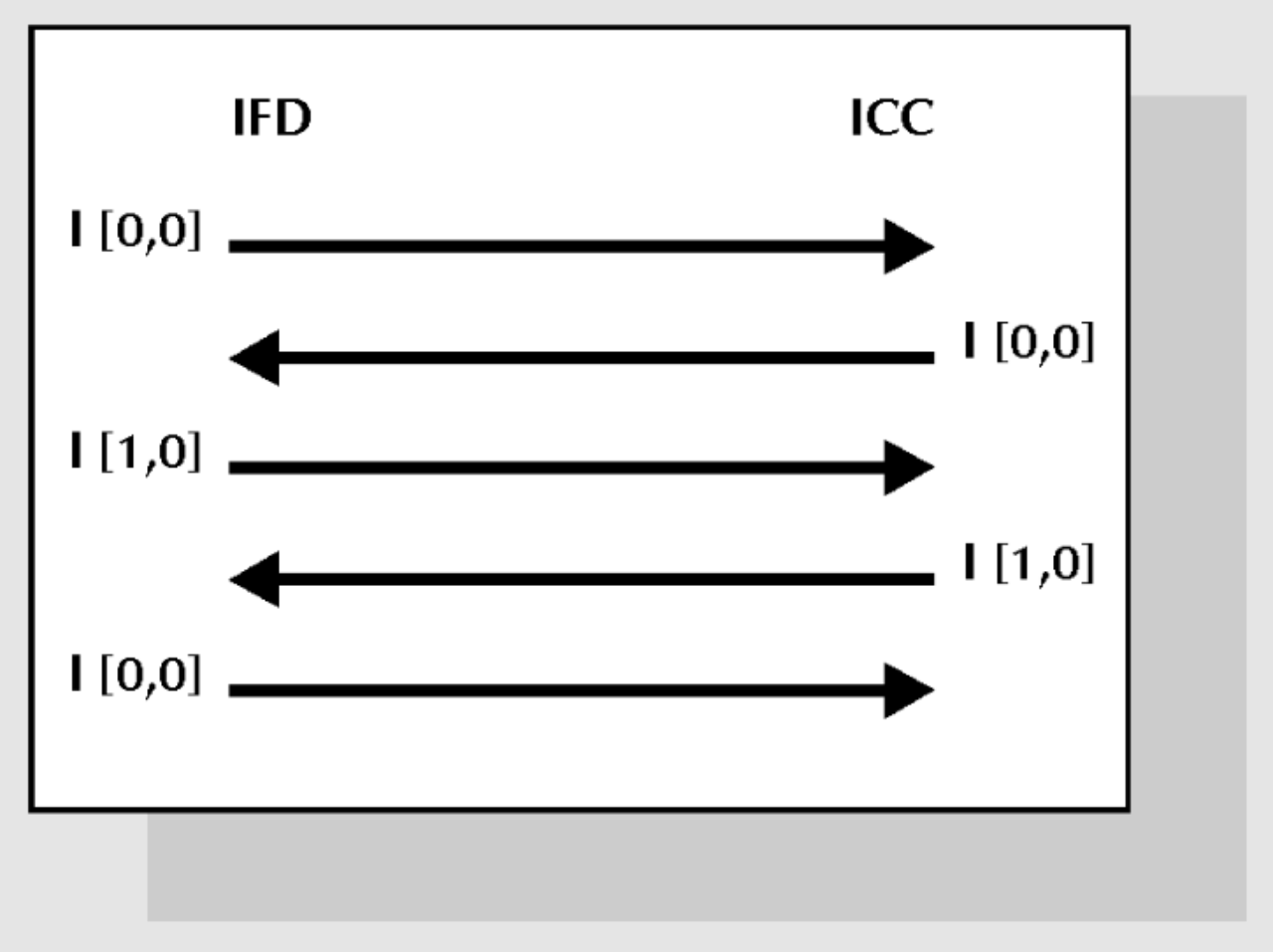
- I Blocks; I (N,M)
- Where N = Sequence number (alternately`0' and `1')
- M = More data bit

The More data bit is set when an additional I block is to follow in the chain

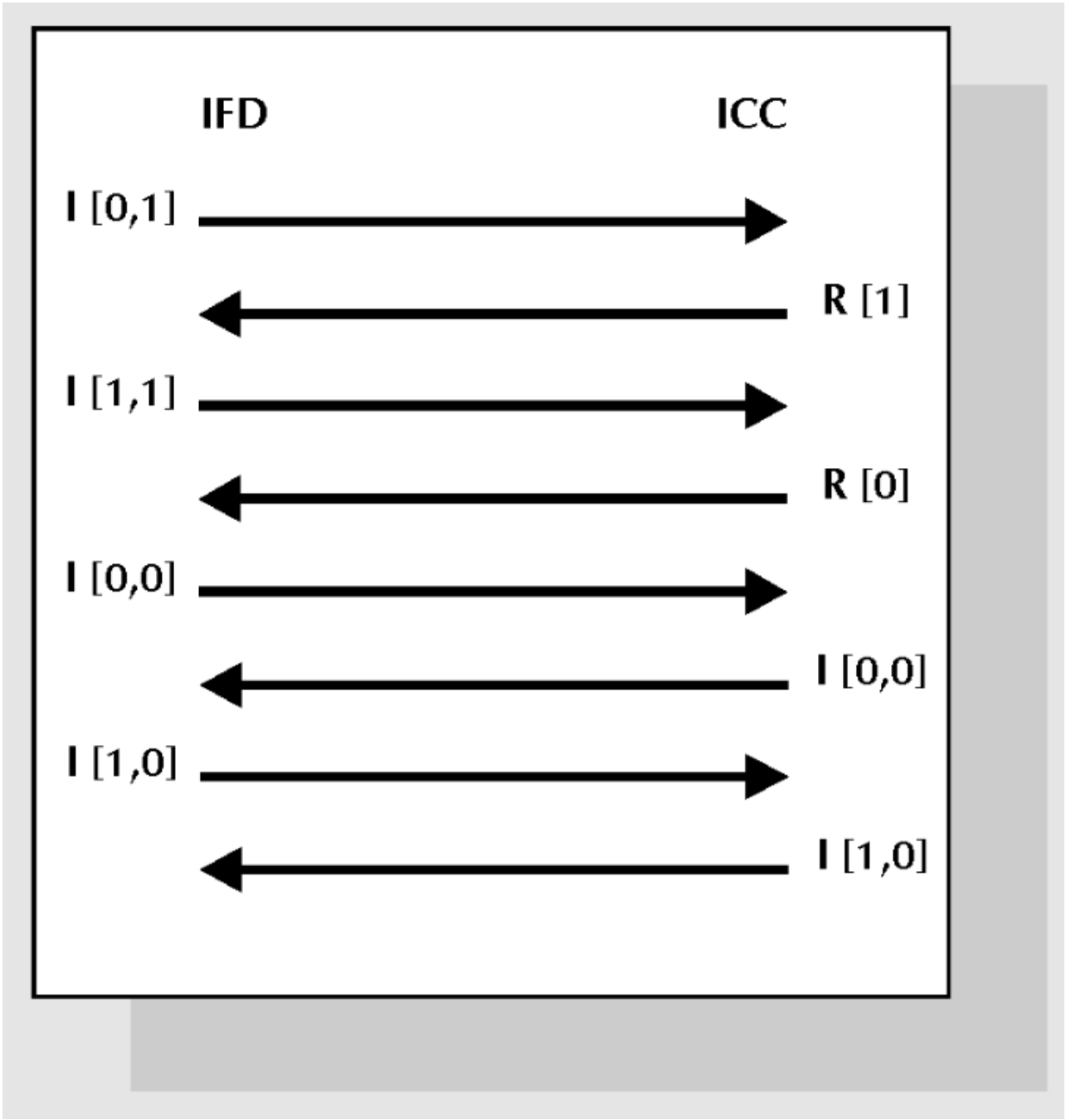
- R Block; R (N)

- Where N = Sequence number of next expected block

The protocol defines that the IFD and the ICC each have the right to transmit in turn where communication commences with transmission of a block by the IFD.

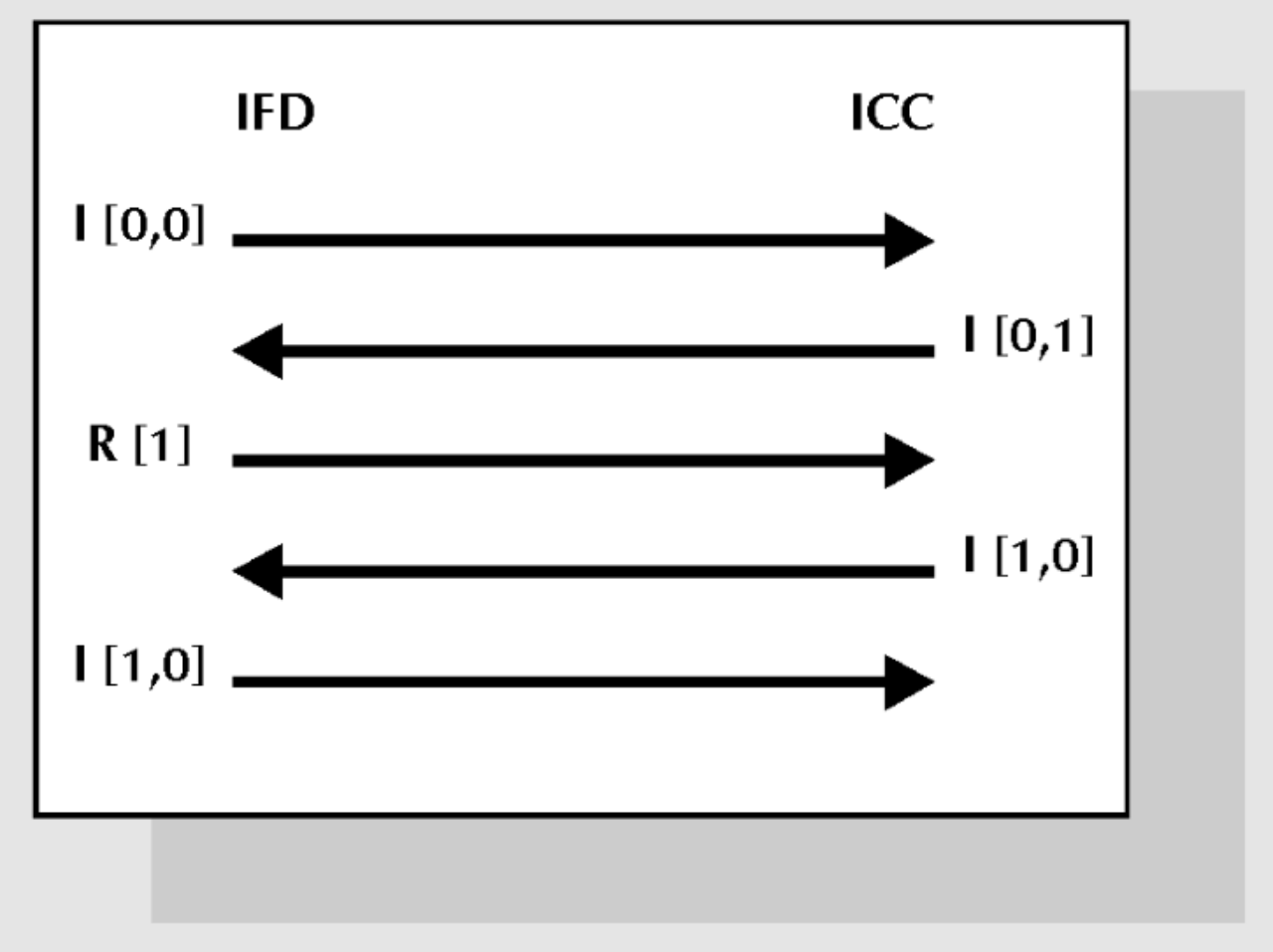


Normal I block transmission



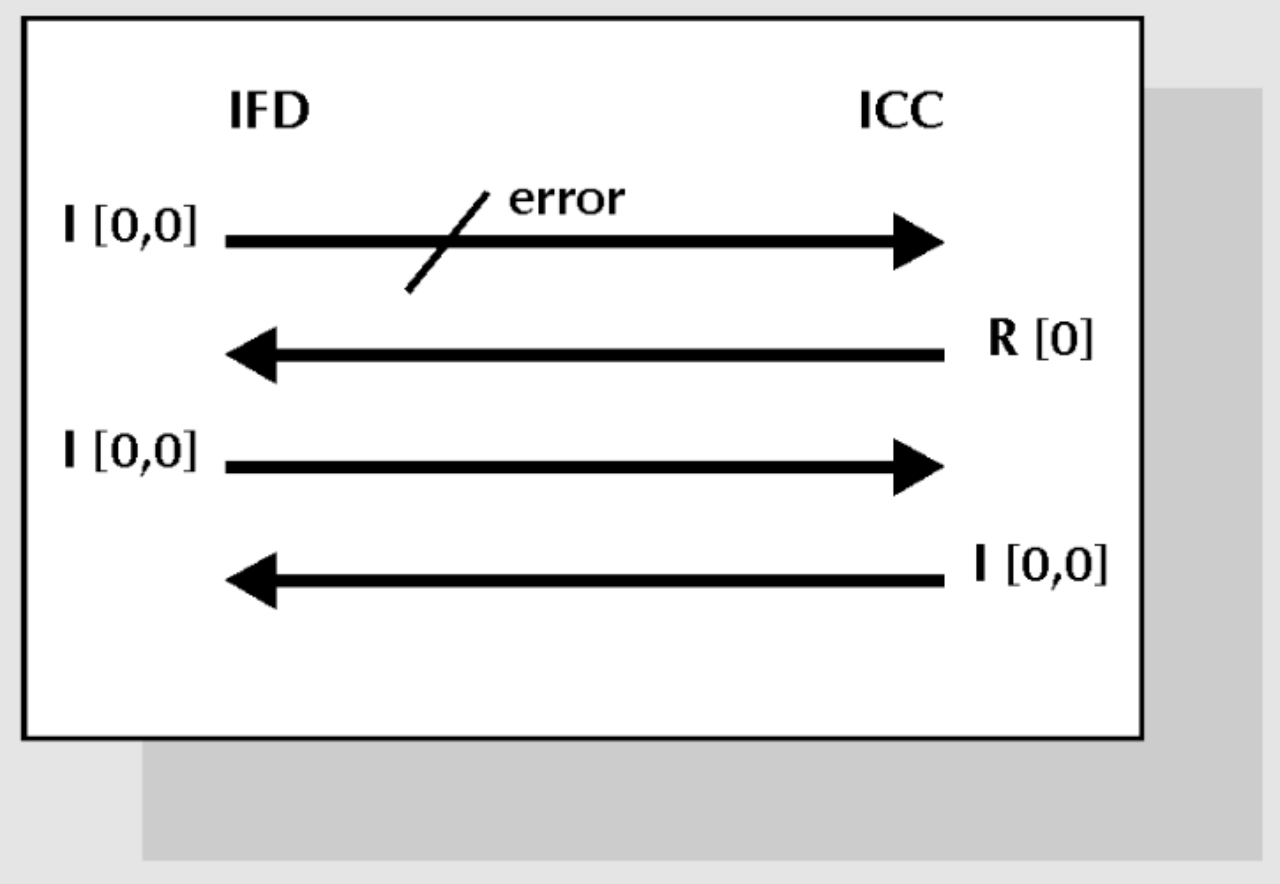
I Block Transmission with chaining

Note that an I block was used by the ICC to acknowledge the last block in the chain sent by the IFD. The ICC may send chained blocks in the same way as shown for the IFD.

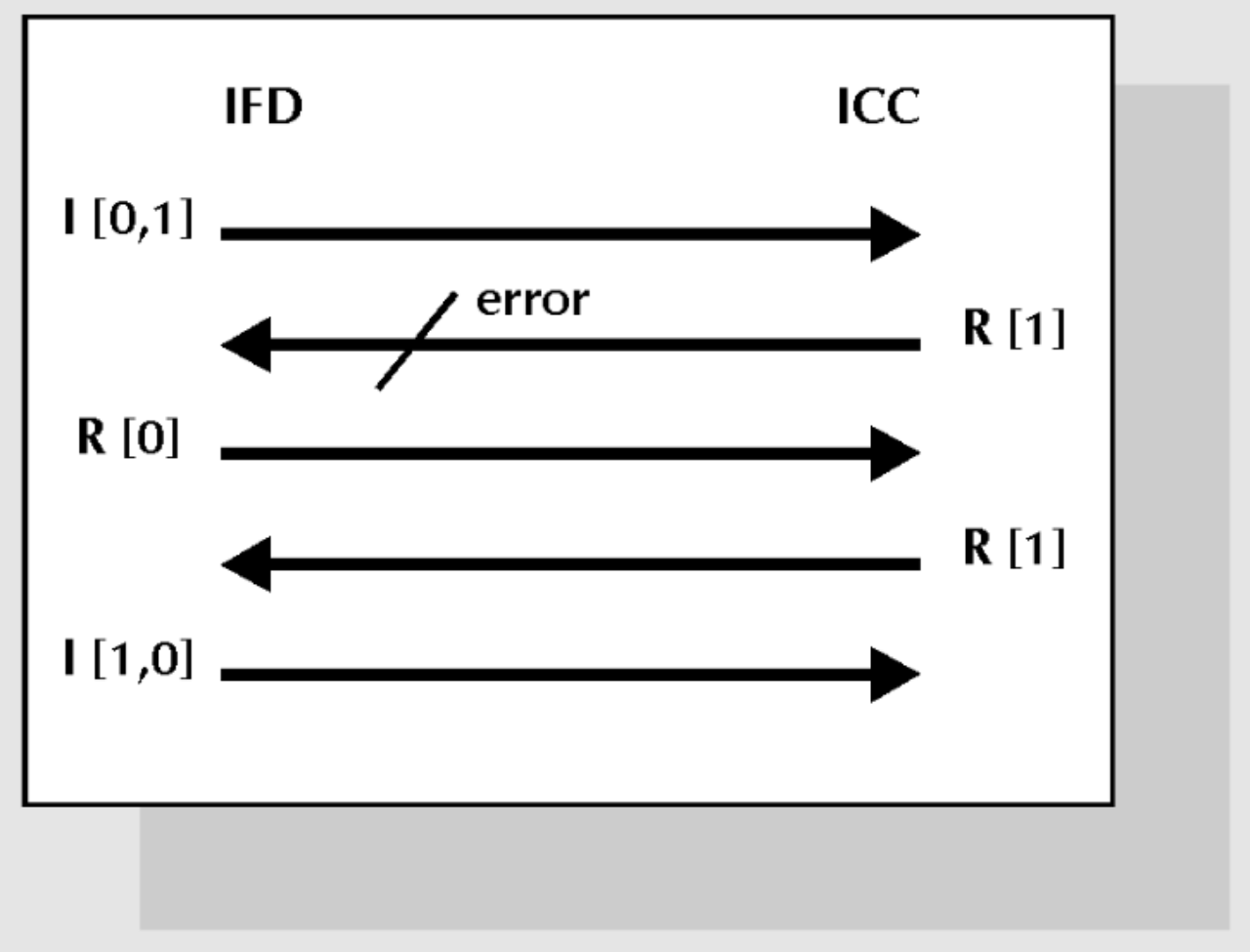


Error handling in I block transmission

Error in I block receipt



Error in I block chain response



In both cases the transmitter is notified to retransmit the block received in error. There are of course a large number of possible error scenarios but they are all based on the simple concepts shown above.

David Everett