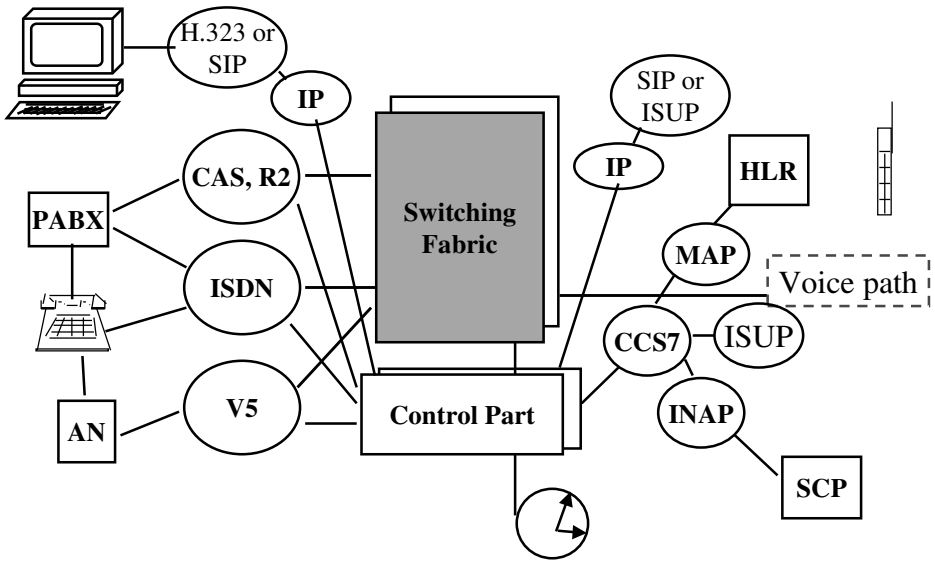
Switching Fabric Technology

Some repetition

Technological issues related to the structure of a switch fabric

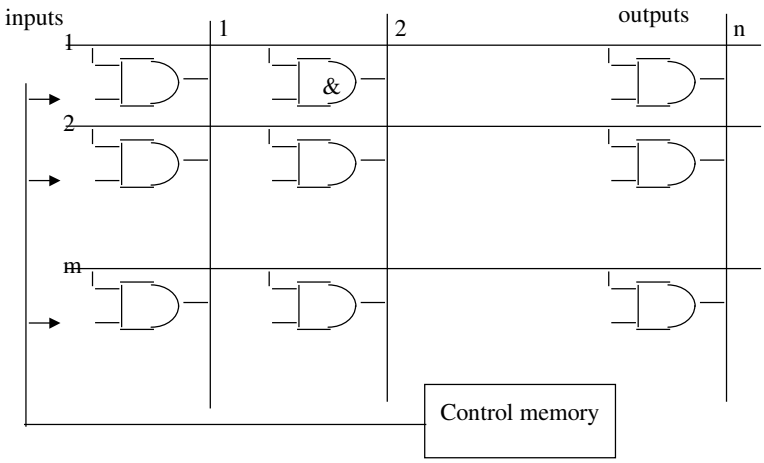
NB: This material is not from the book!

Summary of course scope



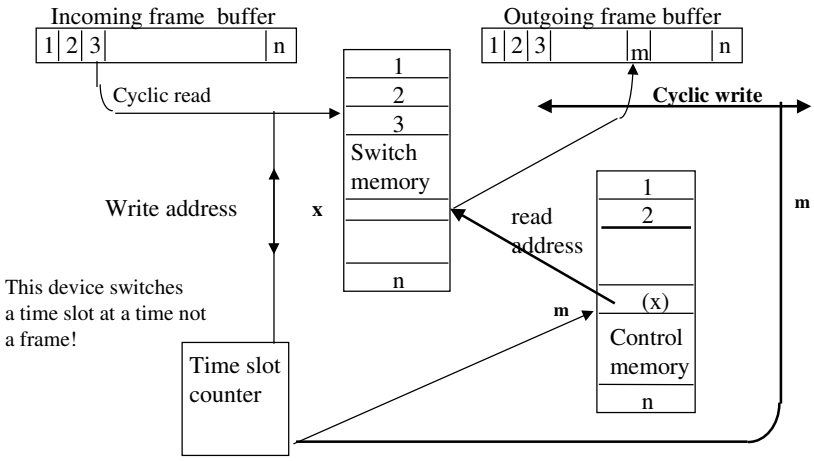
Repetition1: Example implementation of a space switch

A space switch is matrix of cross-points. By opening and closing the cp's information flow may be directed.



Repetition2: A time switch

SM writes are cyclic and controlled by the time slot counter in pace with the multiplexed incoming signal. SM reads are addressed by contents of the CM. Address of CM corresponds to an output slot. CM reads are cyclic.



Properties of time switches

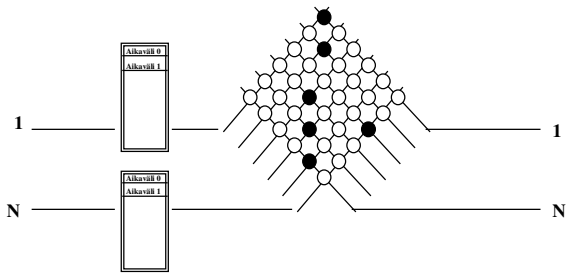
- ✔ **Incoming buffer is fed by the incoming circuit on the "wire" bit rate, outgoing buffer needs to feed the outgoing connection on "wire" bit rate - so, the former needs to be read out on the same speed and in a cycle, the latter needs to be written to on the same speed and in a cycle.**
- ✔ **The number of time slots in a frame = nrof read operations = nrof write operations per frame in the switch memory -> speed of the switch memory is a critical parameter: available speed needs to be made full use of but the same speed determines switch capacity without parallelism.**
- ✔ **It is a good idea of doing Serial-to-parallel and P/S -conversions in the frame buffers - otherwise switch memory speed requirement is multiplied by 8.**
- ✔ **Control memory speed requirement is somewhat above half of the switch memory requirement to allow changing contents i.e. making new thruconnections.**

Two stage switching fabric

- ✔ **Different combinations of Space and Time switches can be used to build a multi-stage switching fabric.**
- ✔ **Possible two-stage combinations of time and space switches are:**
  - › **Time-Time (TT)**
  - › **Time-Space (TS)**
  - › **Space-time (ST)**
  - › **Space-Space (SS)**
- ✔ **TT-fabric does not give any benefit compared to a single stage T -switch.**
- ✔ **SS-switch is not a good idea, because blocking probability is high and no benefit is achieved.**

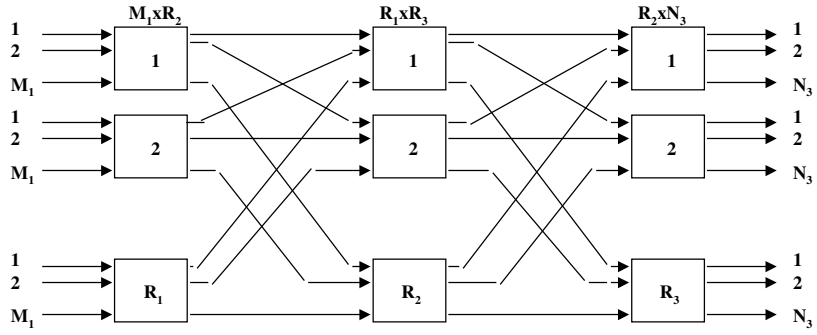
Time- Space switching fabric

- **TS-fabric has low probability of blocking, because the time switch allows rearranging of time slots so that Space switching can be done without blocking.**



Clos Network is a special case of our generic three stage switch fabric

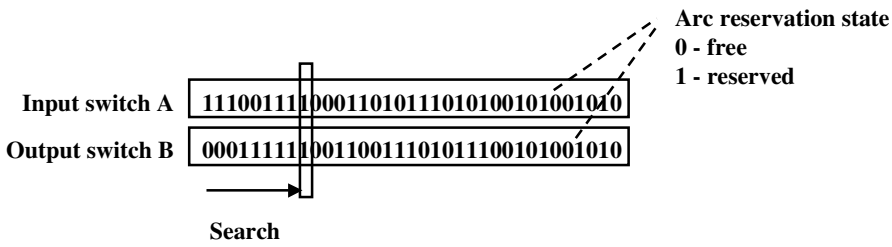
- **In Clos network, each switch block in an earlier stage in connected to each switch block of the following stage with a <u>single</u> link.**



- **Stage 1: $N_1 = R_2$**
- **Stage 2: $M_2 = R_1$ ja $N_2 = R_3$**
- **Stage 3: $M_3 = R_2$**

When input/output signals and the capacity of the fabric are given, the only free variable is $\mathbf{R_2}$

Path search in a Clos network can be done based on *reservation state vectors* reflecting the *use of arcs* between stages



The algorithm can e.g. keep the current position of the search - reflecting the new start point of the search. This leads to even usage of all arcs over time and high likelihood of a fast hit .

Technology 1 - Problems in multi-stage fabrics

- ✓ **Path search is required.**
- ✓ **If fast connection establishment is required, also fast control system is a must**
- ✓ **If control (including path search) is not fast enough, the maximally usable capacity of the fabric is less than theoretic capacity.**
- ✓ **Multicast is not self evident - in fact is complicated matters significantly.**
- ✓ **Multi-slot connections may cause additional problems, if path delay is not constant. Also blocking probability may rise.**

An alternative approach is to take the technological limitations as the starting point

- ✔ **Let us not try to optimize a single parameter (nrof cross-points), but let us look at all limitations at the same time.**
- ✔ **How fast are the available components compared to the wire speeds and slot speeds.**
- ✔ **What is max practical component fan-out.**
- ✔ **How tightly components can be packaged without heat problems due to power consumption.**
- ✔ **How long internal buses are needed in the fabric. Long buses decrease the internal speed in the fabric and also make diagnostics difficult.**
- ✔ **IPR Policy: whether the company wants to use special components or not.**

Faster components drive towards full matrix switches

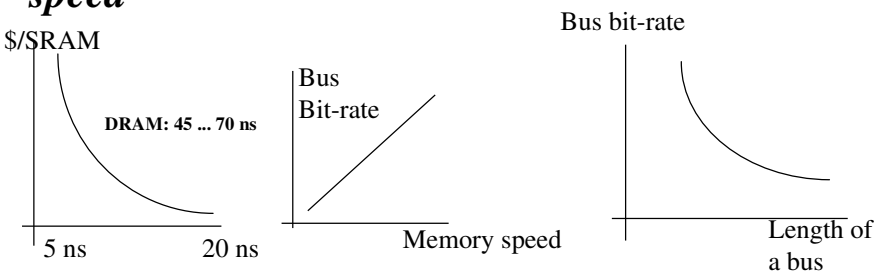
- ✓ **SRAM is faster than DRAM**
- ✓ **Current SRAM technology allows easy implementation e.g. 8k * 2M PCM full matrix - a bigger fabric is hardly needed in Narrow band networks…**
  ==> In NB networks for more than last 10 years the technology trend is to build full matrix switches.
- ✓ **Because customers always want stuff that can hardly be built by engineers and using technology that is not yet properly available (Broadband), multi-stage switches are likely to be needed for a long time to come (cmp. Terabit router).**

Properties of full matrix switches

+ **Strictly non-blocking by definition**
+ **No path search - a thruconnection can always be directly written to the control memory, if the output is free.**
+ **Multicast comes free (1 input --> many outputs)**
+ **Constant delay**
+ **Multi-slot connections are easy**

- **Both switching and control memory size grow in square of nrof inputs/outpts**
- **Broadband --> Necessary memory speeds may not be available -> back to multi-stage**

Good idea: make full use of available memory speed



- ✓ **Price limits the use of fastest components. At time of design, select components that give enough performance, will stay on the market long enough and are not too expensive.**
- ✓ **To make full use of memory speed, buses must be fast. There is a linear relation-ship between the two.**
- ✓ **When required bus speed grows, the practical length of the buses goes down in inverse proportion of faster.**

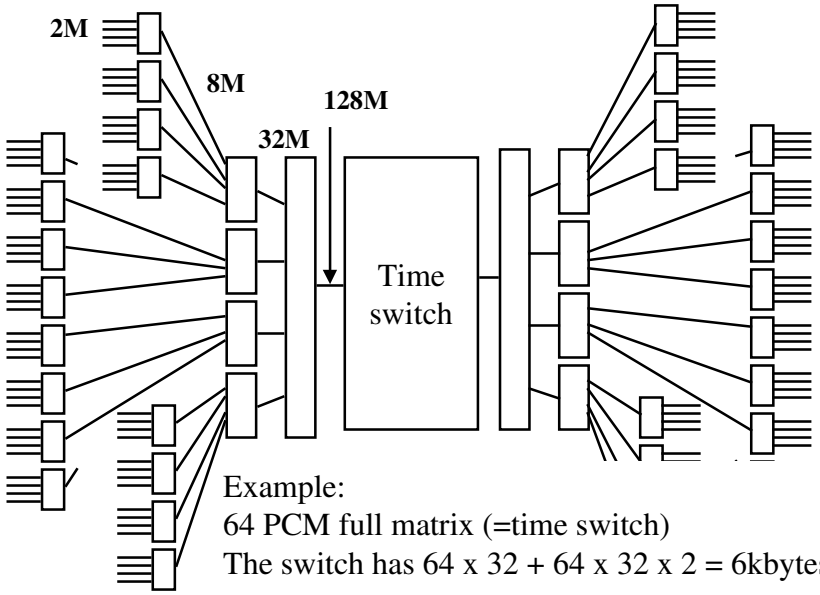
PCM-rates vs. memory speed

Bit-rate	time-slot interval, ns	Bit interval, ns
2M	3906	488
34M	230	29
64 * 2M	61	8
128 *2M	31	4
256 * 2M	15	2

-->64… 256 PCM signals can be written to and read from a SRAM memory continuously on wire speed.

--> Prior to switching, it makes sense to turn the bits in a time-slot into parallel representation

This is easy to implement with SRAM:



Example:
64 PCM full matrix (=time switch)
The switch has 64 x 32 + 64 x 32 x 2 = 6kbytes

Speed requirement for the switch memory in the previous example

Assume: SM is on a single chip:

Size is 64 x 32 octets = 2kbytes

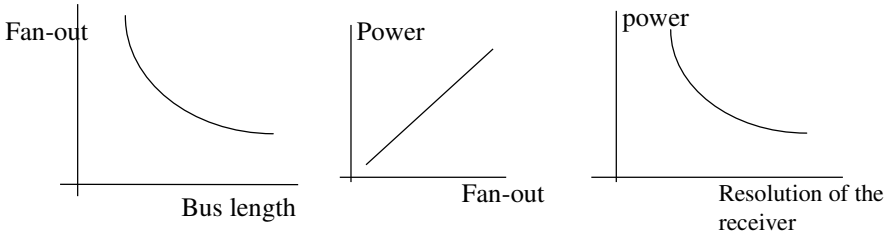
Frame time is 125 μs.

Nrof writes: 2k/125 μs

Nrof reads: 2k/125 μs

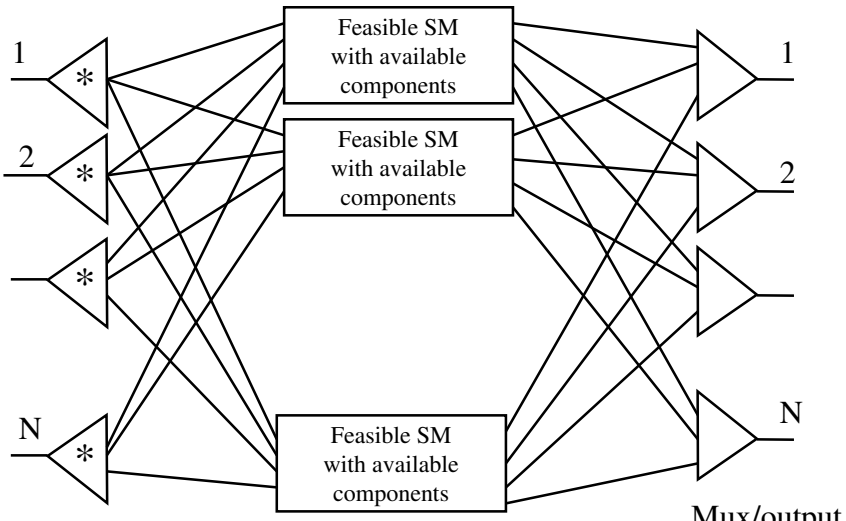
Nrof memory ops: 4k/125 μs = 1/30 ns

Power consuption needs to be kept in check to avoid heat problems



- ✓ **The longer bus needs to be fed by a component of a given nominal fan-out, the smaller is the real fan-out.**
- ✓ **The more components are fed by one output of another, the more power and DC are consumed.**
- ✓ **Power consumption can be reduced with higher resolution receivers (lower voltage components).**

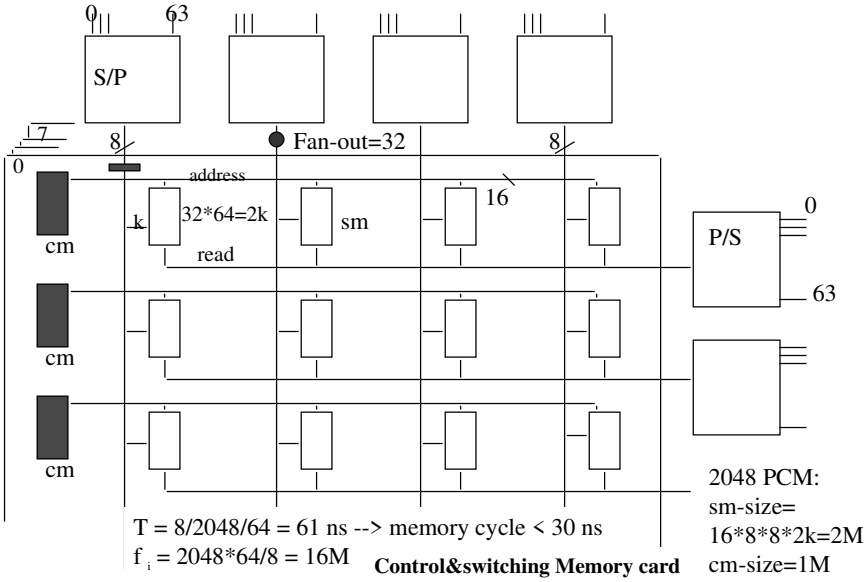
Logical structure of a full matrix switch



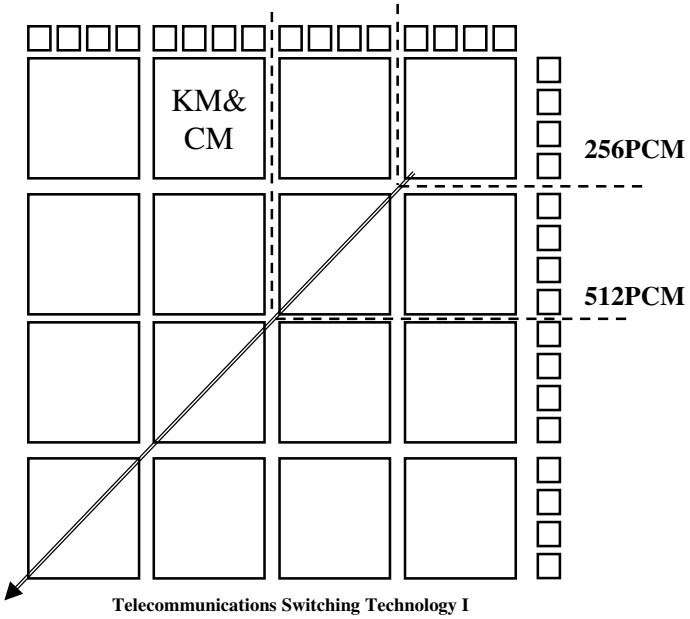
Replication of inputs (multiplexed)

$N = 2^n$

An example of a Matrix switch (DX 200)



Matrix grows in square of nrof inputs



Explanations of the matrix example

- ✓ **S/P - Serial/Parallel converter. Incoming time-slots are turned into parallel representation to reduce the speed on internal buses.**
- ✓ **P/S - prior to outputs signals need to be converted back to serial representation.**
- ✓ **64 PCM S/P + P/S is implemented on one card. This is practical because PCMs are bi-directional.**
- ✓ **One switch block can serve mas 4 S/P+P/S. The Nr is chosen based on required capacity (64, 128, 192 or 256 PCMs).**
- ✓ **One S/P+P/S feeds max 8 parallel switch blocks. Nrof switch blocks is chosen based on the required capacity in the installation (n * 256 PCM's).**
- ✓ **Max size of the example Fabric is 2048 PCM's. Example is from the DX 200 -system.**
- ✓ **Currently, also a bigger matrix (8K PCM's) is available, slightly different SRAMs are needed, principle is similar.**

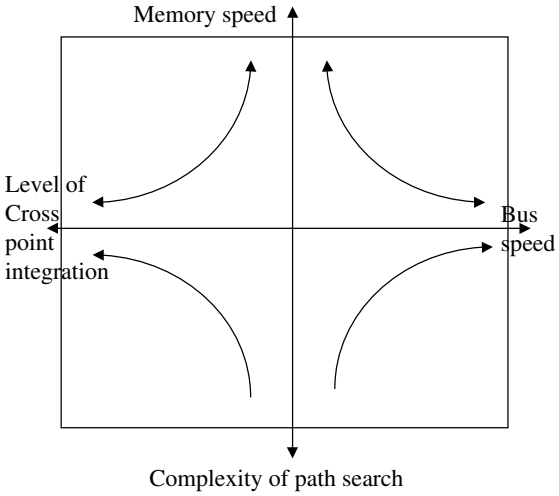
Function of the matrix example

- ✔ **Write from an S/P output goes to all SMs along the vertical bus in all parallel switch blocks making use of the fan-out of the S/P output and the bus buffer.**
- ✔ **Same contents is replicated into max 4x8=32 locations.**
- ✔ **CM contents is used as an address to SM.**
- ✔ **SM(CM(Output-PCM, output-tsl)) = output**
- ✔ **CM(Output-PCM, output-tsl)**
  - › **in a switch block 2 bits of content point to an SM chip and**
  - › **5+6=11 bits point to a memory location on the SM chip.**
  - › **The remaining 3 bits point the switch block (one bit would be enough to say this block/not this block)**

More parameter values of the matrix

- Nrof time slots to be switched during a frame: = 2048 x 32 = 64k
- Switch memory has max 32 x 64 k bytes = 2M
- Only each 32nd memory location is read from SM in a max size switch -> average memory speed requirement is less than the worst case requirement.
- Control memory max size = 2kwords x 4 x 8 x 8 = 4 x 4 x 8 x 8 kbytes = 1 Mbyte.

Technological tradeoffs in a Fabric



When trying to simplify path search and to speed up connection establishment ->
- bus speeds increase-> faster memory is required
- level of integration of the cross-points needs to increase -> faster memory is required

If fast enough memory is not available
--> multi-stage fabrics

Slow path search (BBand)
--> Real capacity may be < theoretical capacity
--> ***minimization of cross-point complexity may be pointless.***