

MONITORY I KARTY GRAFICZNE

Urządzenia techniki komputerowej

GDANSK
styczeń 2003

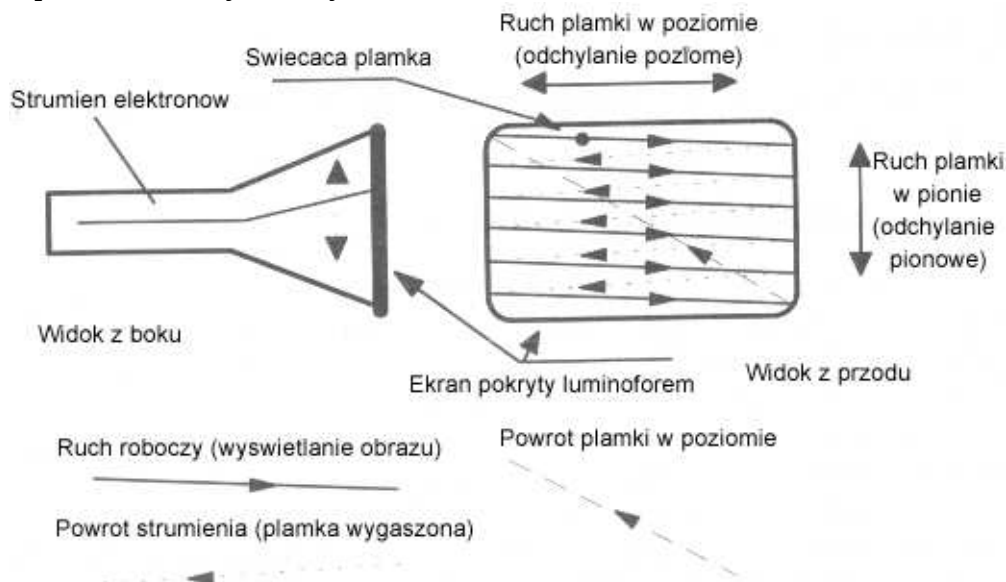


Spis treści

1.1. Zasada działania monitora	2
1.2. Inne rodzaje ekranów.....	6
1.3. Budowa i działanie adapterów graficznych	7
1.3.1. Schemat blokowy prostej karty graficznej.....	8
1.3.2. Praca karty graficznej w trybie tekstowym.....	8
1.3.3. Praca karty graficznej w trybie graficznym.....	10
1.3.4. Karty SVGA	14
1.3.5. Zestawienie własności podstawowych kart graficznych.....	16
1.3.6. Karty akceleratorowe i koprocessorowe.....	17
1.4. Magistrala AGP	18
1.4.1. Właściwości magistrali AGP	19
1.4.2. Wybrane zagadnienia związane z działaniem magistrali AGP	20

1.1. Zasada działania monitora

Monitor jest urządzeniem służącym do wyprowadzania informacji z komputera w postaci obrazów (tekstu bądź grafiki). Zasada tworzenia obrazu na ekranie lampy kineskopowej (ang. *Cathode Ray Tube* - CRT), która jest jednym z zasadniczych elementów monitora, przedstawiona jest na rysunku 1.1.



Rysunek 1.1. Zasada rysowania obrazu na ekranie kineskopu

Ekran kineskopu pokryty jest specjalną substancją zwaną luminoforem. Pod wpływem zogniskowanego strumienia rozpędzonych elektronów luminofor świeci, przy czym jasność tego świecenia zależy od energii (szybkości) elektronów. W przypadku nieruchomego strumienia pozwalałoby to uzyskać na ekranie świecącą plamkę. W celu narysowania obrazu na całej powierzchni obrazu strumień elektronów jest odchylany zarówno w poziomie, co powoduje kreślenie na ekranie pojedynczej linii (przy założeniu stałej jasności świecenia plamki), jak i w pionie, co zapewnia kreślenie kolejnych linii jedna pod drugą. Treść obrazu tworzona jest w ten sposób, że w miarę przesuwania się strumienia elektronów po powierzchni ekranu jego energia jest zmieniana, co powoduje zmiany jasności świecenia poszczególnych punktów luminoforu tworzących punkty wyświetlanego obraz zwane *pikselami* (w naszym przypadku obrazu monochromatycznego).

Kreśląc linię strumień elektronów przesuwają się poziomo z jednej strony ekranu na drugą, na przykład z lewej na prawą, a następnie powinien powrócić ponownie do lewej strony ekranu. Do tworzenia obrazu na ekranie wykorzystywany jest ruch plamki tylko w jedną stronę, przykładowo z lewej na prawą. Wówczas przy powrocie plamki z prawej strony na lewą jest ona *wygaszana*, czyli strumień elektronów jest hamowany na tyle silnie, aby nie powodował świecenia luminoforu. Jednocześnie ruch ten jest znacznie szybszy niż ruch „roboczy” plamki. Z czasem potrzebnym na narysowanie jednej linii i powrotu plamki związany jest parametr zwany *częstotliwością odchylenia poziomego*, będący odwrotnością tego czasu. Inaczej mówiąc, parametr ten określa, ile linii jest kreślonych na ekranie monitora w czasie jednej sekundy.

Na poziomy ruch strumienia elektronów nałożony jest ruch pionowy z góry na dół, wolniejszy od ruchu poziomego, co jak powiedzieliśmy, powoduje kreślenie kolejnych linii jedna pod drugą. W tym przypadku także po dotarciu do dolnego brzegu ekranu strumień

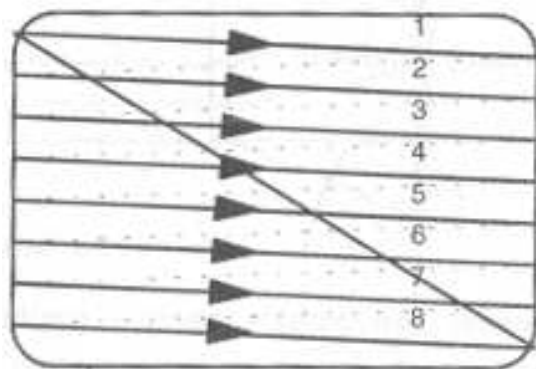
powinien powrócić do górnego brzegu, ruch ten powinien być znacznie szybszy od ruchu w dół i w czasie jego trwania strumień powinien być wygaszony.

Po zapełnieniu całego ekranu treścią obrazu opisany proces jest powtarzany tak często, abyśmy (dzięki bezwładności naszego oka) odnosili wrażenie, że patrzymy na stabilny obraz. Ilość obrazów kreślonych w jednej sekundzie, czyli częstotliwość powtarzania obrazów, zwana jest *częstotliwością odświeżania* obrazu. Jest to jedna z ważniejszych parametrów monitora, decydujący o komforcie pracy, czyli o braku migotania obrazu.

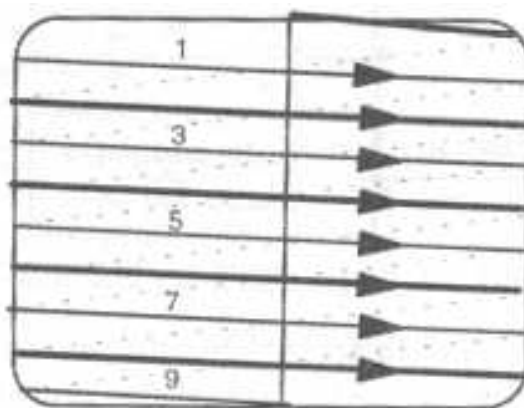
W celu otrzymania na ekranie stabilnego obrazu, każde rozpoczęcie kreślenia zarówno całego nowego obrazu, jak i każdej linii wchodzącej w jego skład, musi rozpoczynać się w takim samym, ściśle określonym momencie. Dlatego też do monitora dostarczane są specjalne sygnały mówiące o tym, kiedy ma się rozpocząć kolejny ruch plamki w poziomie lub pionie. Sygnały te nazywane są odpowiednio *sygnałem synchronizacji odchylenia poziomego* SYNCHRO H (H od ang. *Horizontal*) i *sygnałem synchronizacji odchylenia pionowego* - SYNCHRO V (V ang. *Vertical*). Trzecim sygnałem potrzebnym do uzyskania obrazu jest oczywiście sygnał jasności świecenia plamki, oznaczany często jako sygnał VIDEO.

Wzmacniacz wideo (wizji) wzmacnia sygnał zmieniający energię strumienia elektronów, a więc jasność świecenia plamki. Od szerokości tak zwanego *pasma przenoszenia* tego wzmacniacza zależy, jak szybko możemy zmieniać jasność świecenia plamki, a co za tym idzie, ile szczegółów możemy narysować w danej linii, czyli w poziomie. Parametr ten ma więc wpływ na jeden z bardzo ważnych parametrów monitora, zwany *rozdzielczością*. Rozdzielczość decyduje o ilości szczegółów obrazu, które możemy wyświetlić na ekranie. Rozdzielczość podawana jest jako ilość pikseli możliwych do wyświetlania w poziomie i w pionie. Przykładowo rozdzielczość 640×480 oznacza możliwość wyświetlania 640 pikseli w każdej z 480 linii. Jak powiedzieliśmy, rozdzielczość w poziomie związana jest z pasmem przenoszenia wzmacniacza wideo, natomiast rozdzielczość w pionie wiąże się z ilością linii, które rysujemy na ekranie, a więc z częstotliwością odchylenia poziomego. Na ostrość obrazu, a więc także i na rozdzielczość ma także wpływ średnica plamki (jest ona podawana w milimetrach). Jest to oczywiste, gdy zauważymy, że zbyt duża plamka będzie na przykład powodowała nakładanie się na siebie sąsiadujących linii.

Dla komfortu pracy z monitorem ważny jest brak migotania obrazu, co uzyskujemy powtarzając odpowiednio często wyświetlanie tego samego obrazu na ekranie. Dla uzyskania stabilnego obrazu powinien on być wyświetlany co najmniej 50 razy na sekundę, a dla wyższych rozdzielczości częstotliwość powinna być jeszcze większa. Zwiększanie częstotliwości powtarzania obrazu (czyli częstotliwości odświeżania) wiąże się jednak ze wzrostem ilości informacji przesyłanej do monitora. Dlatego też w pewnych przypadkach stosuje się wybieg pozwalający zmniejszyć tę ilość. Sposób polega na kreśleniu obrazu na ekranie w dwóch etapach, najpierw wszystkie linie parzyste. Inaczej mówiąc, kreślimy na ekranie dwa półobrazy. Powoduje to zwiększenie dwukrotnie częstotliwości powtarzania (pół)obrazu, przy tej samej ilości przesyłanej informacji. Okazuje się jednak, że sposób ten przestaje poprawnie funkcjonować przy wysokich rozdzielczościach obrazu (użytkownicy odczuwają to jednak jako migotanie). Kreślenie obrazu jako jednej całości, linia po linii nazywamy pracą monitora *bez przelotu* (lub wybieraniem kolejnoliniowym, ang. *non-interlaced*), zaś konstruowanie obrazu z dwóch półobrazów zwane jest pracą z *przelotem* (lub wybieraniem międzyliniowym, ang. *interlaced*). Obydwa sposoby wyświetlania obrazu przedstawione są schematycznie na rysunku 1.2.



Praca bez przeplotu

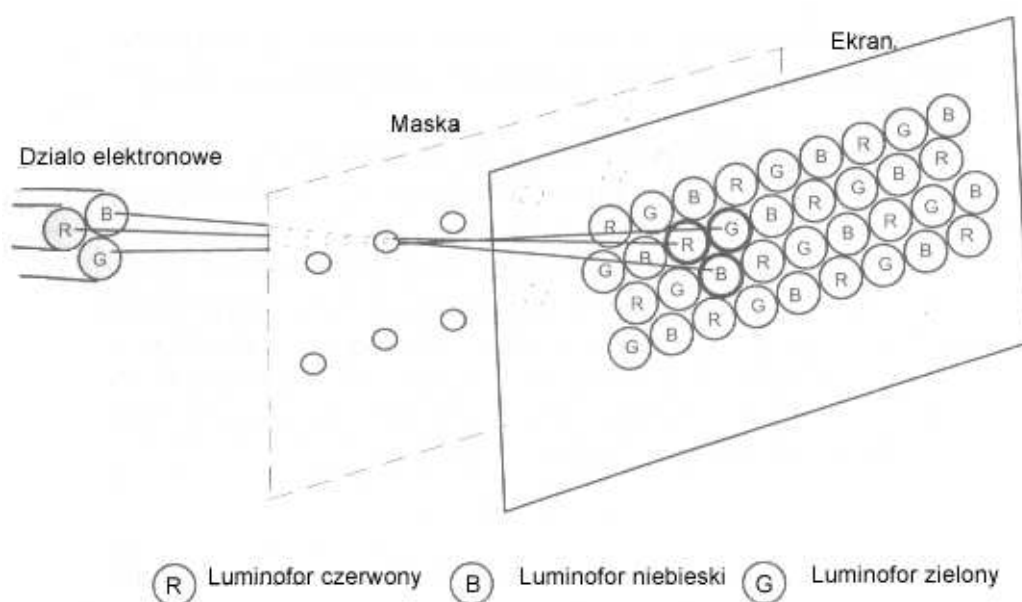


Praca z przeplotem

Rysunek 1.2. Praca monitora bez przeplotu i z przeplotem

Wyświetlanie obrazu kolorowego polega na tworzeniu obrazu z punktów o trzech kolorach: czerwonym (ang. *red*), zielonym (ang. *green*) i niebieskim (ang. *blue*). Kolory czerwony i niebieski są kolorami podstawowymi, kolor zielony zaś jest mieszaniną koloru niebieskiego i żółtego. Mieszając ze sobą w odpowiednich proporcjach wymienione trzy kolory możemy na ekranie uzyskać dowolną barwę.

Ekran kineskopu kolorowego pokryty jest trzema rodzajami luminoforu (tworzącymi punkty bądź paski). Każdy rodzaj luminoforu pobudzany jest do świecenia przez oddzielny strumień elektronów. Schematycznie konstrukcja kineskopu kolorowego pokazana jest na rysunku 1.3.



Rysunek 1.3. Zasada działania kineskopu kolorowego

Jednym z problemów występujących przy budowie tego rodzaju kineskopu jest precyzja wykonania dział elektronowych, układów ogniskujących i odchylających. Nie można bowiem dopuścić, by przykładowo strumień elektronów odpowiedzialny za świecenie koloru niebieskiego trafiał na punkty luminoforu czerwonego. Rozwiązaniem ułatwiającym osiągnięcie prawidłowego stanu jest stosowanie odpowiednio skonstruowanej maski. Oczywiście tego typu kineskop wymaga w miejsce pojedynczego sygnału VIDEO trzech

sygnałów sterujących, zwanych R, G, B, odpowiedzialnych za zmiany jasności świecenia odpowiednio luminoforu czerwonego, zielonego i niebieskiego.

Podsumowując, podstawowymi parametrami monitora są: rozdzielczość, częstotliwość odświeżania, średnica plamki oraz wielkość ekranu. Wielkość ekranu podawana jest w postaci długości jego przekątnej wyrażonej w calach. Jest to jednoznaczne, gdyż stosunek wysokości do szerokości ekranu jest stały i wynosi 4:3. Pomiędzy niektórymi z tych parametrów, taki jak częstotliwość odchyłania poziomego, i pasmem przenoszenia wzmacniacza wideo istnieją dość oczywiste związki. Przy ustalonej rozdzielczości i częstotliwości odświeżania, częstotliwość odchyłania poziomego możemy wyliczyć ze wzoru:

$$f_H = n_H \cdot f_V$$

gdzie:

f_H - częstotliwość odchyłania poziomego

f_V - częstotliwość odświeżania

n_H - rozdzielczość (ilość linii) w poziomie

Wzór jest oczywisty, gdy zauważymy, że w czasie kreślenia jednego obrazu musimy narysować n_H linii, co oznacza, że jedną linię trzeba narysować n_H razy szybciej niż cały obraz.

Zakładając przykładowo pracę monitora z rozdzielczością 640×480 i częstotliwością odświeżania 60 Hz, otrzymujemy częstotliwość odchyłania poziomego równą 28,8 kHz.

Związek szerokości pasma przenoszenia wzmacniacza wideo z rozdzielczością i częstotliwością odświeżania wynika z ilości informacji, którą musimy przesłać przez niego, a dokładnej z maksymalną szybkością zmian sygnału wideo. Przy założeniu, że wyświetlamy na przemian piksele jasne i ciemne, ilość zmian sygnału w ciągu sekundy (a więc i jego częstotliwość) jest równa ilości pikseli, które wyświetlamy w ciągu sekundy. Ilość tę możemy wyliczyć ze wzoru:

$$N = n_H \cdot n_V \cdot f_V$$

gdzie:

N - ilość wyświetlanych pikseli, a więc i częstotliwość sygnału wideo

n_H - rozdzielczość w poziomie

n_V - rozdzielczość pionowa

f_V - częstotliwość odświeżania obrazu.

Oczywiście pasmo przenoszenia wzmacniacza wideo powinno być większe od wyliczonej wartości. Dla wymienionych wyżej parametrów otrzymujemy:

$$f_{\text{Video}} = 640 \cdot 480 \cdot 60 = 18432000$$

i rzeczywiście IBM dla tego trybu podaje wymagane pasmo równe 25,175 MHz.

Ze zmianą rozdzielczości wiąże się (szczególnie w przypadku starszych monitorów) pewne ostrzeżenie. Jak powiedzieliśmy, zwiększenie rozdzielczości „zmusza” generatory odchylenia monitora do pracy z wyższą częstotliwością. Może to spowodować zerwanie synchronizacji drgań oraz przegrzewanie się tych układów, co w konsekwencji może prowadzić do ich uszkodzenia. W przypadku zerwania drgań (brak obrazu na monitorze) należy wyłączyć monitor i powrócić do niższej rozdzielczości (na przykład tryb awaryjny w Windows).

W nowszych monitorach problem ten zwykle nie występuje, monitory te mają bowiem generatory odchyłania, które mogą pracować z różnymi częstotliwościami. Monitory takie określane są wówczas jako monitory *multisync* lub *multiscan*. Jednakże i tutaj nie wolno przekraczać ich parametrów granicznych (jeżeli monitor na to pozwala).

Przykładowe częstotliwości odświeżania i odchyłania poziomego w określonych trybach i przy określonych rozdzielczościach podane są w tabeli 1.1.

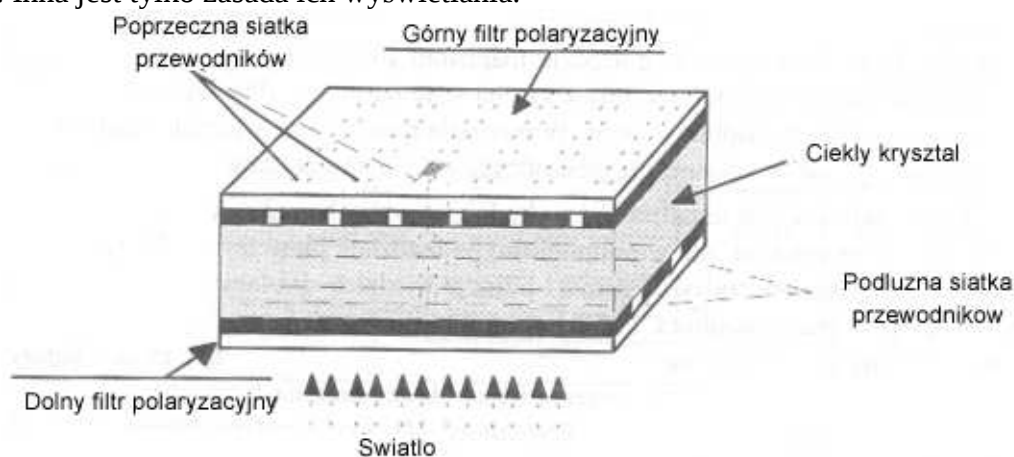
Tabela 1.1. Przykładowe częstotliwości pracy monitorów dla wybranych standardów

Standard	Rozdzielczość	f_v	f_H
EGA	640×350	60 Hz	21,5 kHz
VGA	640×480	60 lub 70 Hz	31,5 kHz
VESA SVGA	800×600	72 Hz	48,1 kHz
VESA SVGA	1024×768	75 Hz	60 kHz
VESA SVGA	1280×1024	75 Hz	80 kHz

Wejście sygnałów wideo może być zarówno analogowe jak i cyfrowe. Problem ten zostanie dokładnie wyjaśniony przy omawianiu adapterów (kart) graficznych.

1.2. Inne rodzaje ekranów

Prócz klasycznych monitorów z kineskopem, stosuje się, szczególnie w komputerach przenośnych, inne rodzaje ekranów, a konkretnie ekrany (wyświetlacze) LCD (ang. *liquid cristal display*) oraz ekrany plazmowe, przy czym te ostatnie są mało popularne. Obydwa typy ekranów są ekranami płaskimi. W jednym i drugim przypadku obraz na ekranie, podobnie jak w przypadku kineskopu, budowany jest z poszczególnych punktów, czyli pikseli. Inna jest tylko zasada ich wyświetlania.



Rysunek 1.4. Zasada działania wyświetlacza LCD

Filtry polaryzacyjne górny i dolny mają płaszczyzny polaryzacji przesunięte względem siebie o 90°. Umieszczony pomiędzy nimi ciekły kryształ powoduje zmianę płaszczyzny polaryzacji światła, przy czym kąt skręcenia tej płaszczyzny zależy od przyłożonego do niego napięcia. W wyświetlaczu ciekłokrystalicznym w taki sposób wybrano ciekły kryształ, aby bez napięcia polaryzującego powodował skręcenie płaszczyzny polaryzacji światła o 90°. Dzięki temu światło może przedostawać się przez wyświetlacz. Przyłożenie napięcia pomiędzy dwa punkty siatki przewodników, podłużnej i poprzecznej, powoduje zmianę kąta skręcenia płaszczyzny polaryzacji światła i przepuszczanie mniejszej ilości światła. W konsekwencji powstaje na wyświetlaczu punkt tym ciemniejszy, im większa

jest wartość przyłożonego napięcia. W przypadku wyświetlaczy kolorowych w odpowiednich punktach umieszczone są filtry o kolorach: czerwonym, zielonym i niebieskim (RGB).

Odmianą ekranu LCD jest wyświetlacz LCD z aktywną matrycą. W tego typu wyświetlaczu każdemu pikselowi odpowiada osobny tranzystor wykonany w tak zwanej technice TFT (tranzystor cienkowarstwowy, ang. *Thin Film Transistor*). Może on przełączać większe prądy niż matryca pasywna i w efekcie uzyskuje się jaśniejszy, bardziej wyraźny obraz.

Inną odmianą płaskich ekranów są ekrany plazmowe. Świeące punkty są tam uzyskiwane w wyniku miejscowej jonizacji gazu. Wymaga to stosunkowo dużego napięcia około 200 V, stąd jak powiedzieliśmy, są one rzadko stosowane.

1.3. Budowa i działanie adapterów graficznych

Terminów adapter i karta można przy pewnym założeniu używać zamiennie. Różnica pomiędzy kartą a adapterem polega na tym, że karta jest urządzeniem wymiennym, które montujemy w gnieździe magistrali rozszerzającej (ang. *slot*), natomiast adapter może być umieszczony także na stałe na płycie głównej. Rozróżnienie to dotyczy wszelkich adapterów i kart. W rozdziale tym używamy terminu karta, gdyż jest on powszechnie stosowane, nie przesadza to jednak o umiejscowieniu urządzenia.

Karta graficzna jest urządzeniem pośredniczącym w komunikacji systemu komputerowego z monitorem. Jak wspomnieliśmy w rozdziale pierwszym, otrzymuje ona informacje o treści obrazu od systemu i po odpowiednich działaniach wytwarza sygnały sterujące pracą monitora. Z biegiem czasu ilość zadań wykonywanych przez kartę graficzną rosła. Sta możemy karty graficzne podzielić na tak zwane *bufory ramki*, które nie uczestniczą w tworzeniu treści obrazu oraz na *karty akceleratorowe*. Wśród tych ostatnich wyróżniamy *akceleratory 2D*, czyli grafiki dwuwymiarowej i *akceleratory grafiki trójwymiarowej 3D*. Zadaniem prostych kart graficznych, czyli buforów ramki, jest wytworzenie sygnałów sterujących monitorem, potrzebnych do uzyskania określonego obrazu. O elementach tego obrazu decyduje zawartość tak zwanej pamięci wideo (o której powiemy w dalszych częściach niniejszego rozdziału). Zawartość ta jest tworzona wyłącznie przez system, a konkretnie przez mikroprocesor, natomiast karta w tym procesie w ogóle nie uczestniczy. W przypadku kart akceleratorowych sytuacja jest nieco odmienna. Mniejsza lub większa część operacji związanych z tworzeniem treści obrazu (inaczej mówiąc, z obliczaniem, jak intensywnie powinny świecić poszczególne piksele składające się na całość obrazu) wykonywana jest przez kartę graficzną (na żądanie systemu). Związane jest to zwykle z potrzebą szybszego wykonania tych operacji i odciążenia procesora. Od strony systemu karta graficzna widziana jest jako układ wejścia / wyjścia współadresowalny z pamięcią operacyjną (zazwyczaj zakres adresów A000:0000 (B000:FFFFh-128 KB). Rejestry konfiguracyjne i sterujące mają zarezerwowane adresy w przestrzeni adresowej izolowanych układów wejścia/wyjścia.

1.3.1. Schemat blokowy prostej karty graficznej

Pracę kart graficznych możemy podzielić na dwa różniące się znacznie tryby: *tekstowy* i *graficzny*. Różnica polega przede wszystkim na sposobie interpretacji zawartości pamięci wideo. W trybie tekstowym zawartość pamięci interpretowana jest jako kody znaków, które należy wyświetlić na ekranie. Zawartość pamięci wideo określa więc w tym przypadku pośrednio, co ma być wyświetlone na ekranie. Informacja o tym, który piksel ma być zapalony, a który zgaszony, pochodzi z generatora znaków.

W trybie graficznym zawartość pamięci wideo jest interpretowana jako bezpośrednie określenie jasności świecenia każdego piksela (przy założeniu, że piksel może być jedynie zgaszony lub zapalony, na każdy piksel przypadałby 1 bit).

Oprócz dostarczania wymienionych sygnałów sterownik monitora adresuje pamięć wideo, matrycę znaków (RAM/ROM znaków) oraz taktuje pracę generatora znaków i rejestru przesuwającego. Oczywiście pełny zestaw funkcji CRTC zależy od jego typu, a zatem i od karty, w której pracuje.

Interfejs magistrali pośredniczy w wymianie informacji karty z CPU (część tych układów jest w istocie zawarta w CRTC). Pamięć wideo zawiera treść obrazu (w formie pośredniej lub bezpośredniej) przeznaczonego do wyświetlania. RAM/ROM znaków zawiera matrycę znaków, czyli informację o sposobie konstruowania znaków z pikseli. Pamięć ta razem z dekoderni atrybutów oraz generatorem znaków współpracuje przy wyświetlaniu znaków w trybie tekstowym. Generatory wyjściowe wytwarzają sygnały o poziomach wymaganych przez monitor, z którym współpracuje karta.

1.3.2. Praca karty graficznej w trybie tekstowym

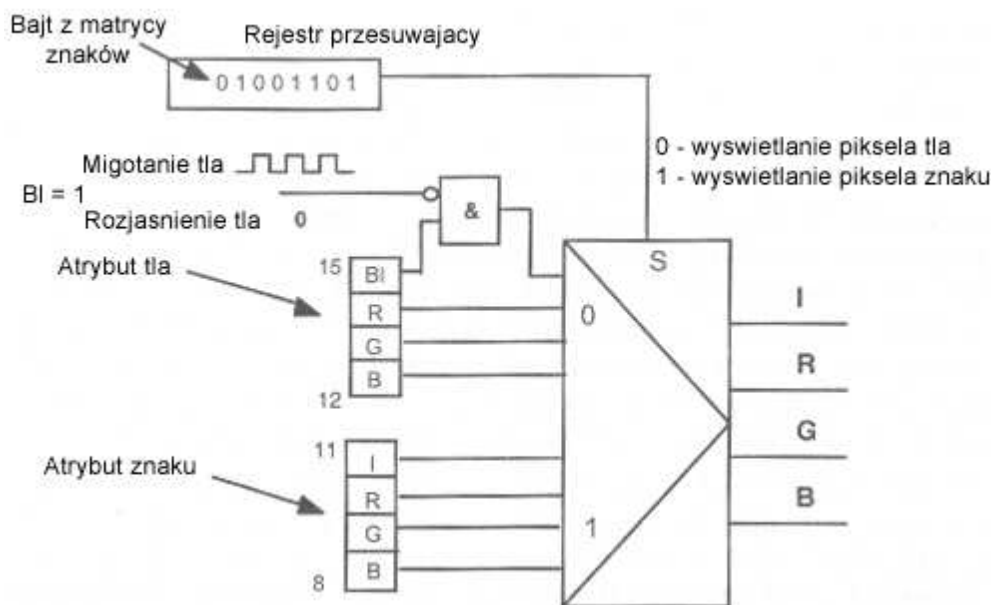
Działanie układu jest następujące: impulsy taktujące przesłanie kolejnych bitów z rejestru przesuwanego podawane są jednocześnie zespołowi liczników, których zadaniem jest śledzenie, który piksel obrazu aktualnie wyświetlany. Licznik L1 liczący modulo m sygnalizuje zakończenie rysowania fragmentu linii należącej do danego znaku (m pikseli na jeden znak) i przejście do rysowania następnego znaku. Impulsy wyjściowe tego licznika zliczane są przez licznik L2. Jego zawartość mówi o tym, który znak w bieżącym wierszu rysujemy. Licznik ten generuje więc część adresu dla pamięci wideo, określając, kod którego znaku należy pobrać. Przepełnienie tego licznika po zliczeniu N_2 impulsów sygnalizuje konieczność zmiany linii, co jest związane z wygenerowaniem impulsu synchronizacji poziomej. Licznik L3 zwiększa swoją zawartość po narysowaniu każdej kolejnej linii. Zliczenie więc n linii przez licznik oznacza zakończenie rysowania bieżącego wiersza i przejście do kreślenia linii należących do wiersza następnego. Przepełnienie licznika powoduje zwiększenie zawartości licznika L4 zliczającego (numerującego wiersze. Zawartość tego licznika stanowi drugą część adresu wybierającego kod określonego znaku w pamięci wideo.

Zawartość pamięci wideo, będąca kodem ASCII aktualnie wyświetlanego znaku, podawana jest do pamięci matrycy znaków, stanowiąc część adresu mówiącą o tym, jakiego znaku konstrukcja będzie aktualnie wyświetlana. Pozostałą część adresu stanowi numer linii aktualnie wyświetlanego znaku pobrany z licznika L3. Powoduje to wybranie określonego bajtu z pamięci matrycy znaków i załadowanie go do rejestru przesuwanego, którego zawartość bit po bicie, zgodnie z taktem zegara przesyłana jest na zewnątrz. Wyjście tego

rejestr steruje jasnością świecenia plamki, co stanowi sygnał wideo (0- pikseli zgaszony, 1 zapalony lub w przypadku wyświetlania koloru 0 - kolor tła, 1 kolor znaku).

Kolory uzyskuje się przy użyciu czterech sygnałów oznaczonych jako I - intensywność, R - czerwony, G -zielony i B - niebieski.

Przy wyświetlaniu znaków w trybie tekstowym w kolorze każdy znak jest reprezentowany w pamięci wideo przez dwa bajty. Pierwszy bajt zawiera kod ASCII wyświetlanego znaku, zaś w drugim bajcie umieszczone są tak zwane atrybuty wyświetlanego znaku. Szczegółowe rozmieszczenie informacji w obydwu bajtach przedstawia rysunek 1.5.



Rysunek 1.5. Wyświetlanie tekstu w trybie znakowym w kolorze

Rysunek 1.5 wyjaśnia użycie bajtu atrybutów w celu uzyskania kolorów znaku i tła. Informacja o konstrukcji znaku załadowana do rejestru przesuwającego nie steruje bezpośrednio jasnością świecenia plamki, lecz jest podawana na wejście sterujące multipleksera. Multiplekser posiada dwa czterobitowe wejścia oraz czterobitowe wyjście. Jeżeli z rejestru przesuwającego podawana jest wartość 1 (co oznacza, że wyświetlamy element znaku) na wyjście multipleksera podawane są bity atrybutów znaku. Cztery bity umożliwiają uzyskanie $2^4 = 16$ różnych kolorów. Zestawienie kolorów w zależności od wartości bitów IRGB przedstawia tabela 1.2.

Tabela 1.2. Kombinacje wartości bitów i odpowiadające im kolory

IRGB	Kolor	IRGB	Kolor
000	Czarny	1000	Ciemnoszary
0001	Niebieski	1001	Jasnoniebieski
0010	Zielony	1010	Jasnozielony
0011	Cyjan	1011	Jasny cyjan
0100	Czerwony	1100	Jasnoczerwony
0101	Purpurowy	1101	Jasnopurpurowy
0110	Brązowy	1110	Żółty
0111	Jasnoszary	1111	Biały

Jeżeli z rejestru przesuwającego podawane jest zero, oznacza to, że wyświetlany jest piksel tła. Wówczas na wyjście multipleksera podawane są bity atrybutów tła. Bity R, G i B podawane są bezpośrednio, natomiast bit B1 (ang. *Blinking* - migotanie) jest podawany za pośrednictwem bramki AND i jego użycie zależy od stanu drugiego wejścia tej bramki. Jeżeli na to wejście podawana jest wartość 0 z tak zwanego rejestru trybu wyświetlania (wchodzącego w skład CRTC), to bit B1 określa rozjaśnienie tła. Jeśli natomiast na wejście to podawany jest przebieg prostokątny (na przemian 1 i 0), to wówczas uzyskujemy migotanie tła (co zresztą nasze oko odbiera jako migotanie znaku).

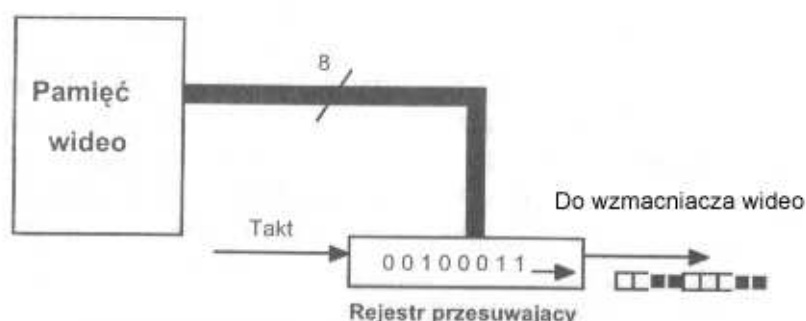
W starszych kartach graficznych matryca znaków umieszczona była w pamięci ROM, co uniemożliwiało jej zmianę. Począwszy od karty EGA matryca ta znajduje się w pamięci RAM.

Na koniec opisu trybu tekstowego wyliczymy pojemność pamięci potrzebną do przechowania zawartości pełnego ekranu w tym trybie, przykładowo dla karty VGA. W jednym z trybów tej karty wyświetlane jest na ekranie 25 wierszy po 80 znaków. Ponieważ w pamięci wideo każdemu znakowi przypisane są dwa bajty, pamięć potrzebna do przechowania treści całego ekranu wynosi:

$$25 \cdot 80 \cdot 2B = 4000B \approx 4 \text{ KB}$$

1.3.3. Praca karty graficznej w trybie graficznym

Informacja z pamięci wideo jest interpretowana jako bezpośrednie określenie jasności świecenia (bądź koloru) pikseli. W sytuacji obrazu czarno-białego (piksel zgaszony lub zapalony) schemat układu wyświetlania jest bardzo prosty. Jest on przedstawiony na rysunku 1.6.



Rysunek 1.6. Tworzenie obrazu czarno-białego w trybie graficznym

Kolejne bajty pobierane z pamięci wideo ładowane są do rejestru przesuwającego. Stamtąd w miarę przesuwania się wiązki elektronów po kineskopie podawane są bit po bicie do układów sterujących jasnością plamki, tworząc kolejne piksele.

Bardziej skomplikowaną sytuację mamy w przypadku wyświetlania obrazu kolorowego. Każdy piksel musi być reprezentowany przez zespół bitów, których ilość zależy od liczby kolorów, których chcemy używać. Ilość bitów, których używamy do reprezentowania każdego piksela musi zapewnić możliwość zakodowania wszystkich używanych kolorów.

Stąd ilość bitów i ilość kolorów muszą spełniać prostą zależność:

$$2^n \geq N \text{ lub inaczej } n \geq \lg_2 N$$

gdzie:

N - ilość używanych kolorów

n- ilość bitów reprezentujących 1 piksel

W starszych kartach graficznych typu EGA czy VGA, ze względu na prostszą budowę układów wyjściowych i stosunkowo małą ilość kolorów, zawartość pamięci nie reprezentowała koloru bezpośrednio. Odpowiednie kombinacje bitów były podawane konwersji za pomocą określonych tablic przekodowujących (ang. *look-up table*), wchodzących w skład układów karty graficznej. Zasadę generowania sygnałów kolorów pokażemy najpierw na prostszym przykładzie karty EGA, a dopiero potem opiszemy analogiczną sytuację w karcie VGA. Ma to dodatkowe uzasadnienie, jako, że część układów zawartych w karcie VGA wynika z możliwości emulowania przez nią karty EGA.

Karta EGA

Pamięć wideo karty EGA dzielona jest na cztery części ważne *płatami* lub *bankami* (ang. *plane*). Z każdego płatu ładowany jest do współpracującego z nim rejestru przesuwającego 1 bajt, co daje w sumie 4 bajty określające wygląd ośmiu pikseli. Kolejne odpowiadające sobie 4 bity pojawiające się na wyjściach rejestrów przesuwających określają kolor wyświetlanego piksela poprzez wybór jednego z szesnastu rejestrów zwanych rejestrami palety kolorów. Pojemność tych rejestrów wynosi 6 bitów, co daje możliwość zakodowania $2^6 = 64$ różnych kolorów. Jednakże bez przeładowywania rejestrów (a więc przeprogramowywania karty) mamy do dyspozycji tylko 16 kolorów, gdyż tyle jest rejestrów palety kolorów (a także tyle różnych możliwości możemy wybrać za pomocą 4 bitów). Stąd mówimy, że karta EGA mogła pracować z maksymalnie szesnastoma kolorami, przy czym były one wybierane z palety 64 kolorów.

Wyjście rejestrów palety kolorów było podawane do monitora w postaci sygnałów R, G, B, IR, IB, IG, przy czym mogły one przyjmować wartości odpowiadające dwóm stanom 0 i 1. Dlatego też wyjście karty EGA określane jest jako wyjście cyfrowe. Gniazdem wyjściowym tej karty było gniazdo typu DB 9. Schematyczny rysunek tego gniazda wraz z określeniami sygnałów występujących na poszczególnych pinach pokazuje rysunek 1.7.



Rysunek 1.7. Schemat gniazda wyjściowego karty EGA

Przyczyną podziału pamięci wideo karty EGA na płaty jest między innymi ilość adresów zarezerwowanych dla kart graficznych w przestrzeni adresowej pamięci. Zakres przydzielonych adresów umożliwia zaadresowanie 128 KB. Dla bogatszych kart EGA oraz dla kart WGA i SVGA jest to ilość zbyt mała, stąd potrzeba podziału pamięci na płaty.

Karta VGA

Karta VGA może pracować w wielu różnych trybach, emulując działanie wcześniejszych kart lub pracując z różnymi rozdzielczościami i liczbami kolorów. Zestawienie trybów pracy karty VGA zawiera tabela 1.3.

Tabela 1.3. Tryby pracy karty VGA

Numer trybu (hex)	Rodzaj trybu	Rozdzielczość / ilość znaków	Ilość kolorów / paleta kolorów	Matryca znaków
0,1	Tekstowy	40×25	16/256K	9×16
2,3	Tekstowy	80×25	16/256K	8×14
4,5	Graficzny	320×200	4/256K	
6	Graficzny	640×200	2/256K	
7	Tekstowy	80×25	mono	9×16
8	Graficzny	160×200	16/256K	
9	Graficzny	320×200	16/256K	
A	Graficzny	640×200	4/256K	
D	Graficzny	320/200	16/256K	
E	Graficzny	640×200	16/256K	
F	Graficzny	640×350	2/256K	
10	Graficzny	640×35016	16/256K	
11	Graficzny	640×480	2/256K	
12	Graficzny	640×480	16/256K	
13	Graficzny	320×200	256/256K	

Przykładowo, dla trybu EGA (tryb 10h) informacja z pamięci ładowania jest do rejestrów przesuwających, skąd za pomocą czterech bitów (tryb 10h jest 16-kolorowy) wybiera 1 z 16 rejestrów palety kolorów. W trybie tym rejestry zawierają swój numer, co powoduje wybranie jednego z pierwszych szesnastu rejestrów w tablicy przekodowującej. Informacja z rejestru palety kolorów podawana jest na tablicę przekodowującą za pośrednictwem multipleksera, przy wartości sygnału S=0 (tryb EGA). Zawartość rejestrów tablicy przekodowującej (18 bitów) decyduje o wyświetlanym kolorze piksela. Ilość kolorów, które możemy wybrać bez przeładowywania rejestrów tablicy przekodowującej, wynosi 16. Należy jednak zauważyć, że w stosunku do oryginalnej karty EGA zmieniła się kolorów, która wynosi 256 K kolorów, co wynika z użycia do kodowania kolorów 18 bitów (takie mamy rejestry w tablicy przekodowującej).

W trybie 13h VGA każdy piksel jest reprezentowany przez 8 bitów, co umożliwia zakodowanie 256 kolorów. Cztery młodsze bity wybierają jeden z szesnastu rejestrów palety kolorów, jednakże w tym trybie każdy z rejestrów zawiera swój zakodowany dwójkowo numer. Oznacza to, że bity docierają niezmienione do tablicy przekodowującej. Następne cztery bity ładowane są do rejestru wyboru kolorów, skąd częściowo bezpośrednio, częściowo zaś za pośrednictwem multipleksera podawane są na tablicę przekodowującą. Ostatecznie 8 bitów wybiera 1 z 256 18-bitowych rejestrów. Na każdą składową koloru poświęconych jest 6 bitów, co daje 64 kombinacje dla jednego koloru. Bite te sterują pracą trzech przetworników a/c, po jednym dla każdej składowej koloru. Łącznie daje to na wyjściu każdego przetwornika 64 różne poziomy napięcia odpowiadające danej składowej koloru, co daje sumarycznie $64 \times 64 \times 64 = 262\,144$ odcienie kolorów. Jak widać, wyjście karty VGA jest wyjściem analogowym (wiele poziomów napięć).

Tabela 1.4. Opis wyprowadzeń gniazda VGA

Nr pinu	Oznaczenie	Opis
1	Red	Sygnał z DAC - kolor czerwony
2	Green	Sygnał z DAC - kolor zielony
3	Blue	Sygnał z DAC - kolor niebieski
4		Zarezerwowane
5	GND	Masa
6	GND	Masa (czerwony)
7	GND	Masa (zielony)
8	GND	Masa (niebieski)
9		Nie podłączone
10	GND	Masa
11		Zarezerwowane
12		Zarezerwowane
13	HSYNC	Synchronizacja odchyłania poziomego
14	VSNC	Synchronizacja odchyłania pionowego
15		Zarezerwowane

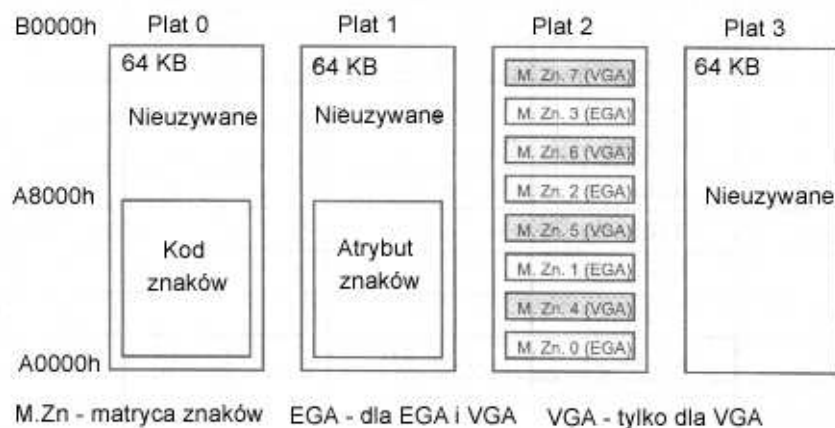
Policzymy teraz ilość pamięci potrzebną dla zapamiętania treści obrazu zajmującego pełny ekran w jednym z wybranych trybów graficznych. Przykładowo dla trybu 12h VGA (16 kolorów przy rozdzielczości 640×480) wymagana pojemność pamięci wynosi:

$$(\text{rozdzielczość pozioma}) \times (\text{rozdzielczość pionowa}) \times (\text{ilość bitów na piksel}) = 640 \times 480 \times 4b = 1\,228\,800 \text{ bitów}$$

czyli (po podzieleniu przez 8) 154 KB. Wielkość tę należy porównać z wymaganiami trybu tekstowego.

Karty VGA, jak wyliczyliśmy, wymagają pamięci wideo o pojemności powyżej 128 KB. Pojemność pamięci karty VGA wynosi standardowo 256 KB. Pamięć ta, podobnie jak dla karty EGA, dzielona jest na cztery banki (płaty) po 64 KB każdy (choć teoretycznie możliwy byłby także podział na 2 banki 128-kilobajtowe). Dla trybów szesnastokolorowych wykorzystanie tych banków jest dość naturalne. Każdy z nich przechowuje bity określające składową jednego koloru. W trybie 256 kolorowym każdy piksel opisywany jest jednym bajtem. Przy rozdzielczości 320×200 wymagana pojemność wynosi $320 \times 200 \times 1B = 64 \text{ KB}$. Informacja o kolejnych pikselach umieszczana jest pod kolejnymi rosnącymi adresami od A0000h do AF9FFh.

Użycie banków pamięci dla trybu tekstowego przedstawia rysunek 1.8.



Rysunek. 1.8. Wykorzystanie banków pamięci wideo dla trybu tekstowego

Jak widzimy, matryce znaków są przechowywane w pamięci RAM, co umożliwia ich zmianę. Zmianę taką najprościej przeprowadzić przy użyciu przerwania 11h VGA BIOS (Interface to the Character Generator).

1.3.4. Karty SVGA

Kartami SVGA nazywa się karty pracujące z rozdzielczościami i/lub ilością kolorów przekraczającymi wielkości podane dla karty VGA. Początkowo karty te projektowane były bez żadnych uzgodnionych standardów, co powodowało wiele kłopotów z kompatybilnością oprogramowania i systemu graficznego. Sytuacja w znacznej mierze została uporządkowana przez komitet o nazwie VESA (ang. *The Video Electronics Standards Association*). Wprowadzono nowe tryby oznaczane liczbami 1xxh dla odróżnienia ich od trybów wprowadzanych przez producentów. Zestawienie nowych trybów SVGA zawiera tabela 1.5.

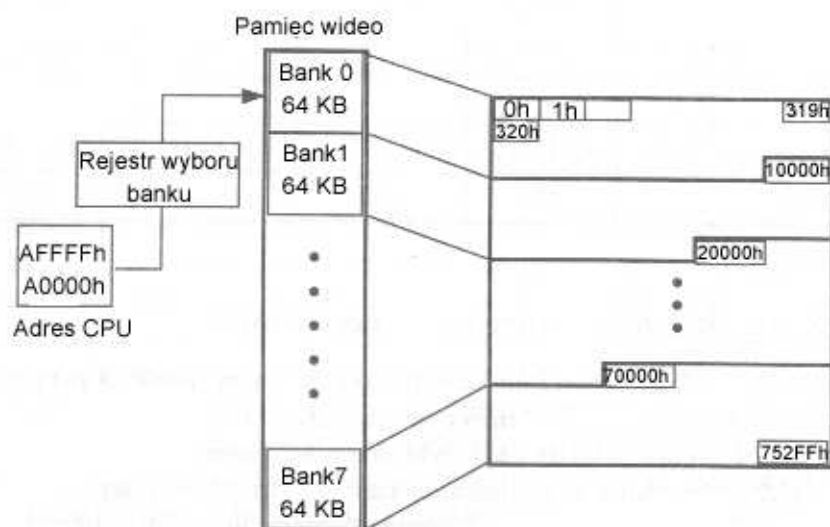
Tabela 1.5. Tryby pracy karty SVGA

Numer trybu (hex)	Rodzaj trybu	Rozdzielczość	Ilość kolorów	Ilość znaków wrs/kol
100	Graficzny	640×400	256	
101	Graficzny	640×480	256	
102	Graficzny	800×600	16	
103	Graficzny	800×600	256	
104	Graficzny	1024×768	16	
105	Graficzny	1024×768	256	
106	Graficzny	1280×1024	16	
107	Graficzny	1280×1024	256	
108	Tekstowy			80 × 60
109	Tekstowy			132 × 25
10A	Tekstowy			132 × 43
10B	Tekstowy			132 × 50
10C	Tekstowy			132 × 60
10D	Graficzny	320×200	32 K	
10E	Graficzny	320×200	64 K	

10F	Graficzny	320×200	16 M	
110	Graficzny	640×480	32 K	
111	Graficzny	640×480	32K	
112	Graficzny	640×480	16 M	
113	Graficzny	800×600	32 K	
114	Graficzny	800×600	64 K	
115	Graficzny	800×600	16 M	
116	Graficzny	1024×768	32 K	
117	Graficzny	1024×768	64 K	
118	Graficzny	1024×768	16 M	
119	Graficzny	1280×1024	32 K	
11A	Graficzny	1280×1024	16 M	
11B	Graficzny	1280×1024	32 K	
11C	Graficzny	1600×1200	32 K	
11 D	Graficzny	1600×1200	64 K	
11E	Graficzny	1600×1200	16 M	

Jak widać z tabeli, do kodowania kolorów używane są 8, 15, 16 lub 24 bity, co daje odpowiednio 256, 32 K, 64 K i 16 M kolorów. Ostatnia z wymienionych możliwości nazywa się trybem True Color. W nowszych kartach stosuje się także kodowanie koloru 32 bitami. Rozwiązanie to jest nieco dyskusyjne, ponieważ oko ludzkie rozróżnia jedynie kilka milionów kolorów. Pewnym uzasadnieniem jest możliwość użycia dodatkowego bajtu do kodowania efektów specjalnych. Ponieważ dla monitorów kolory są kodowane zawsze w systemie RGB (dla drukarek będzie to na przykład CMYK - *Cyjan, Magenta, Yellow, black*), całkowita ilość bitów przeznaczona na kodowanie kolorów dzielona jest na trzy równe części (każda dla jednej składowej koloru). Nie zawsze jest to jednak możliwe, stąd w przypadku 16 bitów na składowe czerwoną i niebieską przypada po 5 bitów, a zieloną 6. Z kolei w przypadku kodowania kolorów 15 bitami z pamięci odczytywane są dwa bajty, a 16 bit jest po prostu nieużywany.

Karty SVGA posiadają pamięci wideo o pojemności 512 KB i większe. Tu także występuje problem adresowania pamięci wideo. Pamięć ta jest najczęściej adresowana przez 64-kilobajtowe okno, przy czym banki pamięci zmieniane są przy użyciu tak zwanego rejestru wyboru banku (ang. *bank select register*). Dane opisujące poszczególne piksele tworzące obraz rozmieszczone są w pamięci wideo liniowo, począwszy od lewego górnego rogu ekranu - najniższy adres, a skończywszy na prawym dolnym - najwyższy adres. Adresowanie i sposób rozmieszczania informacji (adresy pikseli) w pamięci wideo pokazane są na rysunku 1.9 na przykładzie trybu 800×600 z 256 kolorami.



Rysunek 1.9. Obsługa pamięci wideo dla kart SVGA

1.3.5. Zestawienie własności podstawowych kart graficznych

Poniżej zamieszczam tabele, w której zostały zebrane najważniejsze własności podstawowych typów graficznych. Z powodów wyjaśnionych poniżej, do zestawienia tego włączono kartę typu Hercules (ang. *Hercules Graphics Card* - HGC). Część z podanych parametrów zależy od konkretnego wykonania, należy je więc traktować jako wielkości orientacyjne.

Tabela 1.6. Najważniejsze cechy podstawowych typów graficznych

Nazwa parametru	Karta HGC	Karta EGA	Karta VGA	Karta SVGA
Adres pamięci wideo - tekst	B000h	B800h	B000h	B000h
Adres pamięci wideo - grafika	B00h	A000h	A000h	A000h
Rozmiar pamięci wideo	4 KB	64-256 KB	256 KB	≥512 KB
Zakres adresów portów	3B0-3BFh	3C0-3DFh	3B0-3DFh	3B0-3DFh
Matryca znaku	9×14	8×14	9×16	9×16 (typowo)
Maksymalna rozdzielczość	720×348	640×350	640×480	1600×1200 (zależna od karty)
Maksymalna ilość kolorów	Mono	16 z 64	256 z 256 K	16 mln lub 4 mld
Typ sygnałów monitora	Cyfrowy	Cyfrowy	Analogowy	Analogowy
Częstotliwość odchyłania pionowego	50 Hz	60 Hz	50-70 Hz	ok. 50-200 Hz
Częstotliwość	18432 kHz	15,7-21,8 kHz	31,5 kHz	do ok. 90 kHz

odchyłania poziomego				
Pasmo wzmacniacza wideo	16,257 MHz	14,3-16,3 MHz	28 MHz	do ok. 100 MHz
Własny BIOS	Nie	Tak	Tak	Tak

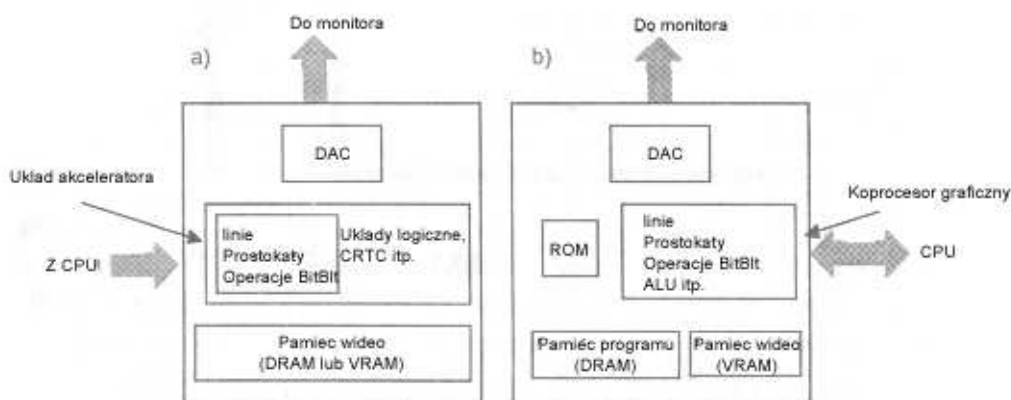
1.3.6. Karty akceleratorowe i koprocesorowe

Opisywane do tej pory karty nie uczestniczyły w żaden sposób w tworzeniu treści obrazu. Jednakże wraz z pojawieniem się *graficznego interfejsu użytkownika* (ang. *Graphical User Interface* - GUI) jak przykładowo w systemach operacyjnych Windows czy OS/2, wymagania w stosunku do kart graficznych wzrosły. Dwoma podstawowymi problemami związanymi z przetwarzaniem dużej ilości informacji graficznych jest:

- przesyłanie dużej ilości informacji poprzez magistralę łączącą procesor główny z pamięcią wideo karty graficznej,
- duże obciążenie procesora głównego obliczeniami związanymi z tworzeniem i zmianami treści obrazu.

Obydwa problemy mogą być w znacznej mierze rozwiązane poprzez przekazanie części „uprawnień” do tworzenia obrazu bezpośrednio karcie graficznej. Zmniejszy to zarówno obciążenie procesora głównego (część wyliczeń wykona karta graficzna), jak i obciążenie magistrali (pamięć wideo znajduje się bezpośrednio na karcie).

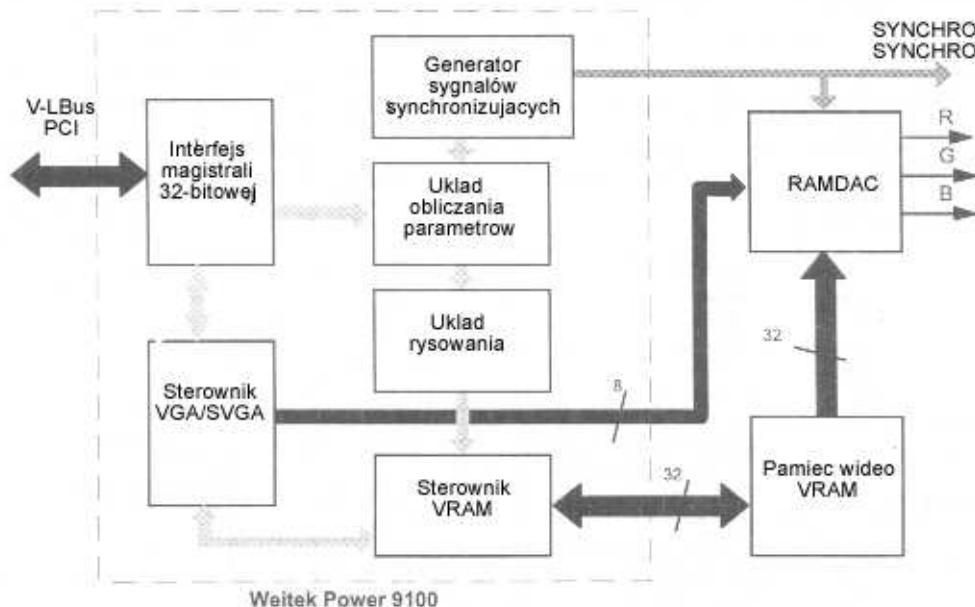
Efektem tych spostrzeżeń jest powstanie kart akceleratorowych i koprocesorowych. Różnica pomiędzy nimi wynika z ich możliwości i źródła, z którymi te możliwości są związane. Dla kart akceleratorowych wynikają one wyłącznie z rozwiązań sprzętowych zastosowanych w kartach, a konkretnie w układzie scalonym akceleratora, stąd ich możliwości są ściśle określone przez ich projektanta. Karty koprocesorowe są wyposażone, jak sama nazwa wskazuje, w procesor oraz w pamięć programu. Możliwości takich kart są więc znacznie większe i zależą także od oprogramowania samej karty. Różnice w budowie kart akceleratorowych i koprocesorowych przedstawione są symbolicznie na rysunku 1.10.



Rysunek 1.10. Uproszczony schemat blokowy: a) karty akceleratorowej, b) karty koprocesorowej

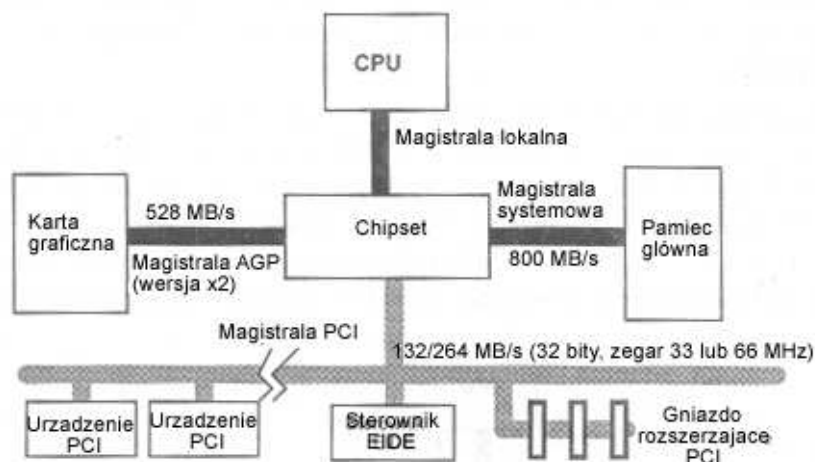
Operacje wykonywane przez tego typu karty, zobrazowane symbolicznie na rysunku 1.10, dotyczą grafiki dwuwymiarowej, związanej, między innymi z graficznym interfejsem użytkownika. Wymaga on częstego wykonywania takich operacji jak tworzenie okien

Jak widać, w układzie zawarte są wszelkie elementy tworzące kartę VGA/SVGA typu bufor ramki (CRTIC, sterownik pamięci wideo, generatory przebiegów synchronizujących itp.) Oprócz wymienionych, układ ten zawiera jednak także bloki odpowiedzialne za realizację operacji graficznych 2D. Do układu obliczania parametrów przekazywane są z CPU polecenia wykonania operacji graficznych (np. rysowania prostokątów, przesuwania obszarów itp.) Na podstawie otrzymanych informacji (charakterystyczne punkty figury np. współrzędne dwóch przeciwległych rogów prostokąta), koloru wypełnienia i tym podobnych, układ ten wylicza odpowiednie parametry, takie jak adresy pamięci wideo i przekazuje je do układu rysowania. Układ ten wypełnia określone obszary pamięci wideo (podanymi przez układ obliczania parametrów) wartościami, tworzącymi za pomocą pikseli żądany obraz.



Kolejnym etapem w rozwoju kart graficznych było opracowanie tak zwanych akceleratorów grafiki trójwymiarowej. Tworzenie tej grafiki wymaga pewnych specyficznych operacji (przykładowo takich jak nakładanie tekstur) oraz bardzo dużej ilości żmudnych obliczeń. Problemy pojawiające się przy używaniu grafiki trójwymiarowej doprowadziły do pojawienia się kart z akceleratorami 3D oraz do opracowania specjalnej magistrali zwanej AGP (ang. *Accelerated Graphics Port*) dedykowanej kartom graficznym, a szczególnie akceleratorom 3D.

Rozwiązaniem obydwóch problemów jest zastosowanie zaproponowanej przez firmę Intel specjalnej magistrali dedykowanej adapterom graficznym, zwanej AGP (ang. *Accelerated Graphics Port*). Schemat blokowy systemu z magistralą AGP przedstawiony jest na rysunku 1.12.



Rysunek 1.12. Schemat blokowy systemu z magistralą AGP

Główne zalety AGP to:

- duża szybkość
- przydzielenie pełnej przepustowości AGP wyłącznie karcie graficznej
- odciążenie magistrali PCI
- umożliwienie wykorzystania (przydzielenia) części pamięci głównej na potrzeby systemu graficznego i przechowywania w niej tekstur.

Z ostatniego punktu wynikają następujące korzyści:

1. Tekstury nie muszą być przed użyciem ładowane do lokalnej pamięci wideo (zysk czasowy).
2. Tekstury mogą być przechowywane w pamięci głównej, co po pierwsze umożliwia użycie tekstur o większych rozmiarach (i rozdzielczościach), a po drugie zmniejsza wymagania co do wielkości pamięci wideo.
3. Zmniejszenie wymagań co do wielkości pamięci wideo znacząco obniża cenę kart graficznych.

1.4.1. Właściwości magistrali AGP

Poniżej przedstawiam podstawowe właściwości magistrali AGP wraz z krótkim opisem ich znaczenia dla systemu.

1. Duża przepustowość. W zależności od wersji (o czym niżej) magistrala AGP oferuje szybkość transmisji danych od 264 MB/s do 1 GB/s. Jak widać, szybkość ta jest porównywalna, a nawet w pewnych wypadkach przekracza transfer PCI. Interfejs dedykowany grafice. Magistrala AGP obsługuje jedynie kartę graficzną (lub oddzielny akcelerator). Cała przepustowość magistrali jest więc przeznaczona dla rej karty.
2. Karta graficzna jest urządzeniem typu master na AGP (inicjatorem w terminologii magistrali PCI). Oznacza to, że inicjacja transmisji zależy wyłącznie od karty graficznej. Urządzeniem docelowym jest chipset zawierający między innymi sterownik pamięci DRAM.
3. Potokowa realizacja transakcji na AGP. Dla AGP, inaczej niż dla PCI, rozdzielono fazę adresowania od fazy transmisji danych. Umożliwia to tworzenie kolejki żądań transakcji (czyli potokową ich realizację).
4. Potokowa realizacja arbitrażu. Zarówno transmisja informacji, jak i zgłoszenie żądania wykonania określonej transakcji wymaga przyznania urządzeniu master dostępu do

magistrali. W celu uniknięcia przestojów (stanów uśpienia magistrali) także arbitraż jest realizowany potokowo).

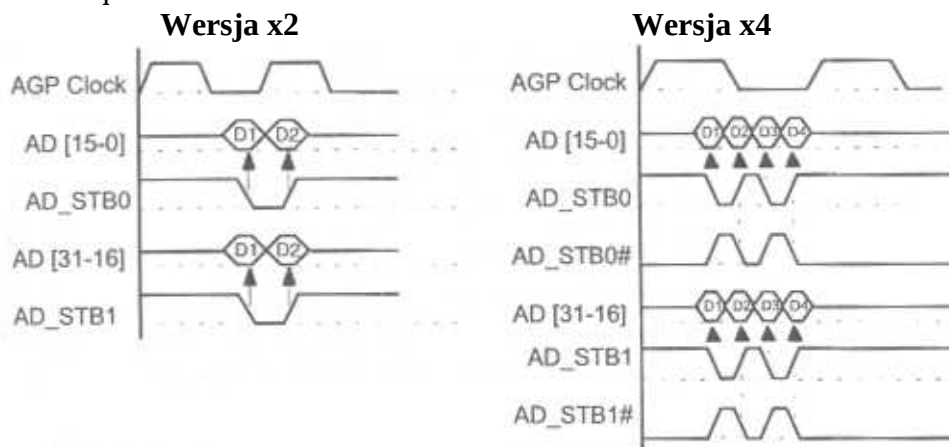
5. Adreoswanie równoległe (ang. *sideband addressing*). Na magistrali AGP możliwa jest inicjacja transakcji nie tylko przy użyciu linii AD I C/BE#, ale także przy użyciu specjalnego 8-bitowego portu oznaczanego jako SBA (ang. *Side-Band Address*). Pozwala on na wykorzystanie linii AD I C/BE# wyłącznie dla transmisji danych.
6. Dynamiczny przydział pamięci. AGP pozwala na przydział części pamięci głównej dla potrzeb systemu graficznego. Jest ona traktowana wówczas jako część pamięci wideo, do której mogą być przykładowo ładowane tekstury. Pozwala to na użycie większych tekstur przy niezwiększonych wymaganiach co do pojemności pamięci na karcie. Przydziału tego dokonuje system operacyjny. W przypadku uruchamiania aplikacji niewymagających tak dużej pamięci wideo możliwa jest dynamiczna dealokacja tej części pamięci i użycie jej dla innych potrzeb (takiej możliwości nie ma dla nieużywanej części pamięci wideo).
7. Przydział pamięci przez AGP. Pamięć główna przydzielona za pośrednictwem AGP na potrzeby adaptera graficznego, jest widziana przez niego jako liniowa przestrzeń adresowa, choć nie musi być ona kontynuacją przestrzeni adresowej pamięci wideo na karcie. Fizycznie, przydzielona pamięć może być nieciągła. Translacja adresu fizycznego na liniowy zapewniana jest przez układy płyty głównej, przy użyciu tablicy umieszczanej w pamięci zwanej tablicą GART (ang. *Graphics Address Remapping Table*). Miejsce umieszczenia tablicy zależy od rozwiązania zastosowanego przez producenta płyty głównej. Przykładowo, sterownik programowy obsługujący taką translację w przypadku systemów Windows 98 czy Windows NT nosi nazwę VGARTD.VXD.
8. Teksturowanie na magistrali AGP. Tekstury w przypadku użycia AGP mogą być przechowywane w pamięci głównej. Akcelerator graficzny wykorzystujący je, odczytuje oraz zapisuje wyniki swej pracy do pamięci głównej za pośrednictwem magistrali AGP. Rezultatem tego jest: skrócenie czasu, obciążenie magistrali PCI i zmniejszenie wymagań co do ilości pamięci wideo.

1.4.2. Wybrane zagadnienia związane z działaniem magistrali AGP

W podpunkcie tym omówię niektóre aspekty działania magistrali AGP wyjaśniające nieco dokładniej pewne jej właściwości.

- Istnieje kilka wersji magistrali AGP o różnej szybkości działania, oznaczanych jako wersje $\times 1$, $\times 2$ i $\times 4$. Istotny jest fakt, że mnożniki te dotyczą wyłącznie szybkości transmisji danych na AGP. Pozostałe operacje są synchronizowane podstawową szybkością zegara AGP (wynoszącą do wersji 2.0 wyłącznie 66 MHz). Sygnały te zwane są często sygnałami pętli zewnętrznej AGP, w odróżnieniu od sygnałów pętli wewnętrznej, związanych bezpośrednio z transmisją informacji. Właśnie sygnały pętli wewnętrznej są synchronizowane zegarem o zwielokrotnionej częstotliwości, zależnej od wersji AGP. Sposób traktowania transmisji danych uzyskuje się poprzez transmisję na obydwu zboczach sygnałów AD_STB $\times$ (mają one tę samą częstotliwość co AGP Clock). Dla wersji $\times 4$ sygnały AD_STB0 i AD_STB1 są sygnałami o podwojonej w stosunku do AGP Clock częstotliwości. Transmisja jest taktowana zboczem odpowiadającym sygnałom AD_STB $\times$ i AD_STB $\times$ #.
- Istnieje dwie wersje napięciowe magistrali AGP: 3,3 V i 1,5 V. Wartości te dotyczą napięcia na wyprowadzeniu oznaczonym Vddq, decydującym o poziomach logicznych magistrali (nie jest to napięcie zasilania układów na karcie). Magistrala w wersji $\times 1$ i $\times 2$ może pracować w dowolnym z tych napięć, magistrala w wersji $\times 4$ używa tylko tych wartości 1,5V. Przykładowo, dla środowiska 3,3 V poziomy logiczne na wejściach układów wynoszą: niski od - 0,5 V do 0,99 V, wysoki od 1,65 V do 2,8 V. Gniazda mogą

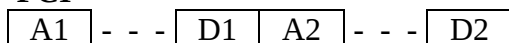
być gniazdami przeznaczonymi dla konkretnego napięcia lub uniwersalnymi. W przypadku określonych napięć gniazda wyposażone są w klucze zabezpieczające przed włożeniem niewłaściwej karty (dla gniazda 3,3 V klucz przy kontaktach nr 22-25, dla gniazd 1,5 V kontakty 42-45). W przypadku gniazda uniwersalnego układy logiczne płyty głównej muszą rozpoznać typ karty przy użyciu kontaktu TYPDET# i zapewnić stosowne napięcie Vddq.



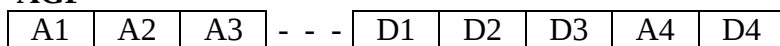
Rysunek 1.13. Sygnały taktujące transmisję danych na magistrali AGP

- Na magistrali AGP możliwe są dwa sposoby uzyskania dostępu do pamięci. Pierwszy z nich to użycie adresowania równoległego, wymienionego we właściwościach AGP. Drugim sposobem jest zgłoszenie dostępu i podanie adresu przy użyciu linii AD i C/BE#. W tym wypadku działanie AGP jest nieco odmienne niż PCI. Magistrala AGP pozwala na podanie kolejnych adresów przed zakończeniem realizacji wcześniejszych żądań transmisji. Różnica ta jest zilustrowana na rysunku 1.14.

PCI



AGP



Rysunek 1.14. Adresowanie równoległe na magistrali AGP

- W tabelicy 1.7 zestawione są rodzaje cykli wykonywanych przez AGP. Jak widzimy transmisje dla odczytu mogą być „zwykłe” lub „długie” (ang. *long*). Ilość transmitowanych bajtów jest określona w trakcie żądania transmisji w polu długości transmisji, zgodnie z tabelą 1.7 a (transmisje zwykłe) lub 1.7 b (transmisje długie). Dostęp do pamięci głównej przez AGP dotyczy zawsze argumentów będących czterosłowem (64 bitami, inaczej mówiąc stosowane adresy są zawsze podzielone przez 8. Także ilość transmitowanych bajtów jest podzielona przez 8.

Transmisja może mieć niski priorytet (ang. *low priority*) lub wysoki (ang. *high priority*). Transmisje o wysokim priorytecie mają gwarantowany czas realizacji. Kolejność realizacji transmisji ulega zmianie tylko w wypadku niemożności dotrzymania tego czasu dla transmisji priorytecie. Cykl o nazwie Dual Address Cycle realizowany jest w wypadku konieczności zastosowania adresu 64-bitowego.

Tabela 1.7. Kody operacji magistrali AGP

Kod operacji	Rodzaj operacji
--------------	-----------------

0000	Low-Priority Read
0001	High-Priority Read
0100	Low-Priority Write
0101	High-Priority Write
1000	Long Low-Priority Read
1001	Long High-Priority Read
1101	Dual Address Cycle

Tabela 1.7 a Kody długości transmisji “zwykłych

Kod długości	Liczba bajtów
000	8
001	16
010	24
011	32
100	40
101	48
110	56
111	64

Tabela 1.7 b Kody długości transmisji „długich”

Kod długości	Liczba bajtów
000	32
001	64
010	96
011	128
100	160
101	192
110	224
111	256