Wojna o chipy: jak krzem stał się walutą władzy

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje globalną rywalizację o technologię półprzewodników, która stała się kluczową walutą władzy w geopolityce, obronności i ekonomii. Przedstawia reguły oceny siły technologicznej, takie jak reguła łańcucha funkcjonalnego, podkreślając znaczenie każdego etapu – od projektowania (EDA) po pakowanie (HBM, CoWoS) i testowanie. Omówione są praktyczne zastosowania kluczowych technologii chipów (CPU, GPU, akceleratory AI, układy rad-hard) w kontekście wojskowym, kosmicznym i cywilnym. Artykuł zgłębia również geopolityczne aspekty kontroli nad łańcuchem dostaw, wpływ embarg technologicznych i sankcji na rozwój państw, ilustrując to przykładami Huawei, SMIC i Rosji. Zrozumienie tych mechanizmów jest kluczowe dla oceny strategicznej pozycji państw w erze cyfrowej.

This article examines the global competition for semiconductor technology, which has become a key currency of power in geopolitics, defense, and economics. It presents principles for assessing technological power, such as the functional chain rule, emphasizing the importance of each stage—from design (EDA) to packaging (HBM, CoWoS) and testing. It discusses the practical applications of key chip technologies (CPUs, GPUs, AI accelerators, rad-hard chips) in military, space, and civilian contexts. The article also explores the geopolitical aspects of supply chain control and the impact of technology embargoes and sanctions on state development, illustrating these mechanisms with examples from Huawei, SMIC, and Russia. Understanding these mechanisms is crucial for assessing the strategic position of states in the digital age.

Aby precyzyjnie ocenić technologiczną siłę, warto posłużyć się dwiema roboczymi regułami, które działają niczym analityczna suwmiarka. Pierwsza to reguła łańcucha funkcjonalnego. Przewagi w kosmosie i wojsku nie zdobywa się bowiem „najlepszym chipem”, lecz najstabilniejszym połączeniem projektowania (EDA i IP), wytwarzania (węzeł litograficzny i kontrola procesu), pakowania (HBM i interposery), testowania (metrologia i MTBF) oraz

serwisu (kalibracje i części), a wszystko to spięte jest ludzką kompetencją. Przerwiesz dowolne ogniwo i reszta stanie się droższą wersją przeciętności. Druga to reguła ekonomii decyzji. Jeśli dana technologia skraca czas potrzebny do podjęcia wiarygodnej decyzji na brzegu systemu – w samolocie, na orbicie, na morzu czy w pojeździe lądowym – to nawet przy wyższej cenie jednostkowej okazuje się tańsza w skali całej kampanii. Redukuje bowiem łączny czas, zużycie amunicji, energii i, co nie mniej ważne, ryzyko polityczne. Taka perspektywa pozwala zupełnie inaczej spojrzeć na skuteczność embarg. Embargo nie jest guzikiem „on/off" dla postępu, tylko nastawem zaworu na rurociągu uczenia się. Jeśli odcina dostęp do maszyn EUV, zaawansowanej metrologii i oprogramowania EDA, to spowalnia akumulację doświadczenia procesowego – tej części przewagi, której nie da się odtworzyć samą lekturą dokumentacji. Jeśli obejmuje serwis, działa jak ogranicznik mocy w silniku: masz sprzęt, ale nie jesteś w stanie trwale...

SŁOWA KLUCZOWE

wojna o chipy | krzem waluta władzy | technologia półprzewodników | embarga technologiczne | łańcuch funkcjonalny | ekonomia decyzji | CPU | GPU | HBM (High Bandwidth Memory) | CoWoS (Chip-on-Wafer-on-Substrate) | EDA (Electronic Design Automation) | RISC-V | akceleratory AI | układy rad-hard | litografia | pakowanie chipów

Półprzewodniki: fundamenty przewagi technologicznej

trzymać go w optymalnej kondycji. Jeżeli zaś dotyczy gotowych akceleratorów AI, to nie zabija samych algorytmów, ale przenosi ciężar z „uczymy się dużo i szybko" na „uczymy się dłużej i mniej". Dla państw o ogromnym rynku wewnętrznym to wciąż jest wykonalne, lecz w wyścigu z czasem fundamentalnie zmienia klasyfikację zawodnika. Druga strona odpowiada na to surowcami, co tylko potwierdza, że gra toczy się na polu, gdzie każdy ruch ma swoje lustrzane odbicie.

CPU, GPU, NPU, FPGA: strategiczne zastosowania

Zacznijmy od prozaicznego, a jednak wciąż królewskiego CPU, czyli uniwersalnego procesora sterującego. W epoce misji Apollo komputer pokładowy AGC, zbudowany na krzemowych układach Fairchilda i obsługiwany przez legendarny panel DSKY, był nie tylko mózgiem nawigacji, lecz także poligonem

doświadczalnym dla industrializacji mikroelektroniki. NASA, jako wymagający klient, wymusiła powtarzalność jakości, a laboratorium MIT zaprojektowało oprogramowanie na miarę ówczesnej fizyki. Z perspektywy organizacji był to przełom: państwowy program zmusił prywatny łańcuch dostaw do porzucenia rzemiosła na rzecz żelaznej dyscypliny procesu. Bez tego przymusu dojrzewanie technologii krzemowej trwałoby znacznie dłużej.

Kiedy jednak wchodzimy w przestrzeń kosmiczną na poważnie, sama uniwersalność przestaje wystarczać. Na scenę wkracza kategoria „rad-hard", czyli układy scalone utwardzone na promieniowanie. RAD750 firmy BAE Systems, potomek rodziny PowerPC, to ikona tej filozofii. Jego zegary rzędu 110–200 MHz i wydajność kilkuset MIPS mogą brzmieć skromnie w erze akceleratorów AI, ale w kosmosie liczy się determinizm i wieloletnia dostępność części, a nie bicie rekordów. W tej skali zaufanie do procesu produkcyjnego i rygorystycznej kwalifikacji jest ważniejsze niż jakikolwiek benchmark. Kosmos po raz kolejny uczy pokory.

Tymczasem na lądzie i w chmurze obliczeniowej rządzi dziś GPU, czyli wyspecjalizowany procesor do przetwarzania równoległego. Jego znaczenie w wojsku i kosmosie eksplodowało, gdy rozpoznanie obrazowe i fuzja danych sensorycznych przestały mieścić się w budżecie czasowym klasycznego CPU. Rzecz jednak w tym, że współczesna „moc bojowa" GPU nie jest już tylko funkcją zaawansowania litografii, lecz przede wszystkim przepustowości do pamięci HBM oraz dostępności zaawansowanego pakowania. HBM, czyli High Bandwidth Memory, to pamięć 3D z niezwykle szeroką magistralą, która w najnowszych generacjach przesuwa wąskie gardło dokładnie tam, gdzie najbardziej boli operatorów systemów rozpoznawczych: w dostępie do danych. To nie jest subtelna ewolucja; to rewolucja w architekturze.

Skoro pamięć staje się arterią, pakowanie staje się chirurgią naczyniową. Technologia CoWoS, czyli Chip-on-Wafer-on-Substrate, to dziś prawdziwy drugi front walki o wydajność. Bez dostępu do mocy produkcyjnych CoWoS nawet najdoskonalszy układ logiczny pozostaje bezużyteczną krzemową płytką w magazynie. Przez cały 2024 rok TSMC gorączkowo rozbudowywało swoje zdolności w tym zakresie, ponieważ stało się to kluczowym wąskim gardłem dla dostaw akceleratorów AI. Projektanci rezerwują przepustowość z wielomiesięcznym wyprzedzeniem, niczym kiedyś czas antenowy w telewizji. Ekonomia mocy obliczeniowej zaczęła przypominać energetykę szczytową, gdzie liczy się nie tylko ilość, ale i dostępność w kluczowym momencie.

Na tym tle naturalnie wyrasta NPU, czyli akcelerator skrojony na miarę operacji tensorowych w sieciach neuronowych, oraz FPGA, czyli programowalna „glina" sprzętowa, używana w łączności i prototypowaniu. W zastosowaniach wojskowych i kosmicznych FPGA ma dwa kuszące atuty: możliwość aktualizacji swojej funkcji już po starcie misji oraz relatywnie dobrą odporność na trudne warunki. Mimo to najwięcej finansowego tlenu łapią dziś GPU i NPU. Nowoczesne systemy rozpoznania, od analizy obrazu po radar z

syntetyczną aperturą, wymagają nie tyle prostych obliczeń, co nieprzerwanego strumieniowania danych. Stąd właśnie triumf architektury „HBM blisko krzemu” i agresywny rozwój technologii łączących chipy.

Wróćmy na chwilę do systemów sterowania i naprowadzania, bo tu widać najdłuższą linię ewolucji. Pociski Tomahawk z lat 80. i 90., z algorytmami TERCOM i DSMAC, uczyły się dopasowywać obraz terenu z radaru do wgranych wcześniej map satelitarnych. Dzisiejsze systemy robią to samo, ale szybciej, z większą precyzją i na podstawie fuzji danych z wielu sensorów. Co ważniejsze, przenoszą część wnioskowania na samą platformę. Ma to oczywisty sens operacyjny: jeśli na trasie przelotu zmienia się pogoda lub oświetlenie, system musi mieć zdolność adaptacji „tu i teraz”. Kiedyś robił to potężny komputer w centrum planowania misji; dziś w dużej mierze robią to układy scalone tuż przy antenie. Zysk jest prosty: krótsza pętla decyzyjna i mniejsza zależność od zawodnych łączy danych.

Oczywiście żaden z tych światów nie istnieje w geopolitycznej próżni. Otwarta architektura RISC-V wdarła się do debaty strategicznej, ponieważ usuwa część „opłat licencyjnych” – nie tyle finansowych, co politycznych. W Chinach widać było wyraźny wzrost zainteresowania RISC-V, a nawet zapowiedzi wsparcia platformy CUDA dla systemów z procesorami opartymi na tej architekturze. To ruch, który można czytać dwojako: inżyniersko, bo heterogeniczność jest przyszłością, oraz politycznie, jako próbę utrzymania kompatybilności oprogramowania tam, gdzie dostęp do sprzętu podlega restrykcjom. Nie oznacza to, że x86 i Arm znikną z mapy. Znaczy to tylko tyle, że te mapy przestały być jednokolorowe.

Aby ułożyć to wszystko w głowie praktyka, sprowadźmy rzecz do żywych definicji. CPU jest dyrygentem orkiestry, świetnym w logice sterowania i zadaniach o przewidywalnej strukturze. GPU to sekcja smyczkowa i perkusja naraz, która przerabia ogromne bloki danych, pod warunkiem że ma szerokie gardło do pamięci; bez HBM dusi się jak alpinista bez tlenu. NPU to waltornia wyspecjalizowana w graniu na tensorach: oszczędna w energii, ale mało elastyczna. FPGA to „lutownica w czasie rzeczywistym”, pozwalająca zmieniać układ połączeń na orbicie lub w polu. Z kolei układy analogowe i radiowe to zmysły całego systemu. Bez nich na wejściu mamy szum, a nie sygnał. I wreszcie pamięć – to nie magazyn, lecz układ krążenia. Im bliżej serca, tym mniejsze straty.

CoWoS i HBM: klucz do mocy obliczeniowej

Na lądzie i w chmurze rządzi dziś GPU, czyli wyspecjalizowany procesor do przetwarzania równoległego. W wojsku i kosmosie jego znaczenie eksplodowało z chwilą, gdy rozpoznanie obrazowe i fuzja danych przestały mieścić się w budżecie czasowym klasycznego CPU. Rzecz jednak w tym, że współczesna „moc bojowa” procesora graficznego nie jest już funkcją samego węzła litograficznego, lecz przepustowości do pamięci HBM oraz dostępności zaawansowanego pakowania. HBM, czyli High Bandwidth Memory, to trójwymiarowa pamięć o niezwykle szerokiej magistrali, która w wersji HBM3E osiąga blisko 9,6 Gb/s na

pin. Nowa generacja HBM4 podwaja tę magistralę do 2048 bitów i celuje w przepustowość rzędu 1,6–2,0 TB/s na stos, przesuwając wąskie gardło dokładnie tam, gdzie boli operatorów ISR: w dostępie do danych. Ten ruch nie jest subtelny; to rewolucja czasoprzestrzenna.

Skoro pamięć staje się arterią, pakowanie staje się chirurgią naczyniową. CoWoS, czyli technologia łączenia wielu chipów na jednej podstawie, i jej warianty to dziś prawdziwy drugi front „litografii", bez którego nawet najdoskonalszy układ logiczny pozostaje bezużyteczny w magazynie. TSMC przez cały 2024 rok i później intensywnie zwiększa moce produkcyjne CoWoS, ponieważ to właśnie one stały się wąskim gardłem dla dostaw akceleratorów. Projektanci rezerwują przepustowość z wyprzedzeniem, niczym niegdyś czas antenowy. Jak ujął to Jensen Huang, zmienia się nie tyle ogólne „zapotrzebowanie na pakowanie", ile precyzyjna potrzeba zwiększenia wolumenu konkretnego wariantu, by utrzymać krzywą wzrostu mocy. Jeśli ktoś chce zobaczyć gospodarcze znaczenie inżynierii opakowań, niech spojrzy na cenniki rezerwacji tych slotów. Ekonomia mocy obliczeniowej zaczęła przypominać energetykę szczytową.

Apollo, Minuteman, Tomahawk: chipy w służbie wojska

Program Apollo nie opierał się wyłącznie na brutalnej sile silników F-1 i odwadze astronautów. W jego nerwowym centrum pracował Apollo Guidance Computer (AGC), pierwszy w historii cyfrowy komputer sterujący załogowym lotem kosmicznym. Stworzony w MIT Instrumentation Laboratory, stanowił akt inżynieryjnej brawury: jego architekturę oparto na układach scalonych Fairchilda, które w latach 60. były technologią niestabilną i horrendalnie drogą. Każdy AGC zawierał około 5 tysięcy bramek logicznych, taktowanych zegarem 2 MHz, i korzystał z pamięci ferrytowej. W porównaniu z ENIAC-iem, kolosem o 18 tysiącach lamp próżniowych zajmującym salę balową, była to miniatura. Ale to właśnie ta miniatura wyniosła ludzkość na Księżyc, co inżynierowie NASA kwitowali zwięźle: „Bez układów scalonych nie byłoby Apollo".

Rakiety balistyczne Minuteman II były pierwszym systemem wojskowym, który zaczął pożerać półprzewodniki na skalę przemysłową. Gigantyczny kontrakt Pentagonu z Texas Instruments wywindował popyt z dziesiątek do setek tysięcy układów, zmuszając firmę do transformacji. Z laboratorium badawczego TI musiało przeistoczyć się w przemysłowego giganta zdolnego do masowej, niezawodnej produkcji. W tym momencie układ scalony przestał być ciekawostką naukowców, a stał się kluczowym elementem narodowej strategii odstraszania nuklearnego. Krzem stał się równie ważny jak pluton.

Wojna w Wietnamie okazała się poligonem doświadczalnym dla bomb Paveway, pierwszej generacji broni precyzyjnego rażenia. Ich zasada działania była prosta: prymitywny czujnik laserowy, garść tranzystorów i miniaturowy komputer pozwalały zmienić bezwładny lot „głupiej bomby" w precyzyjny manewr śledzenia odbitego od celu promienia. Efekt był porażający – zasięg błędu spadł ze 130 metrów do zaledwie kilku.

Koszt jednostkowy Paveway był astronomiczny, lecz w chłodnej logice całkowitego kosztu posiadania jedna precyzyjna bomba okazywała się tańsza niż dywanowy nalot kilkudziesięciu B-52.

Lata 80. i 90. należały do pocisków Tomahawk, które stały się medialną ikoną amerykańskiej „inteligentnej siły". Ich skuteczność była owocem miniaturyzacji. Komputer pokładowy, radarowy wysokościomierz oraz systemy TERCOM i DSMAC, pozwalające pociskowi porównywać rzeźbę terenu i cyfrowy obraz celu z wgraną mapą, byłyby niemożliwe bez gęsto upakowanych chipów. Dzięki nim Tomahawk potrafił lecieć setki kilometrów na pułapie kilkudziesięciu metrów i uderzać z chirurgiczną precyzją. Transmitowane przez CNN podczas wojny w Zatoce Perskiej w 1991 roku, stały się globalnym symbolem nowej, technologicznej ery wojen.

Embarga na chipy: mechanizm i złożoność

Embargo na chipy nie przypomina blokady na węgiel czy stal; jest raczej precyzyjnym cięciem w układzie nerwowym globalnej technologii. Półprzewodniki to produkty o ekstremalnie wydłużonym i kruchym łańcuchu wartości, w którym każdy element – od maszyn litograficznych, przez oprogramowanie do projektowania EDA, po specjalistyczne surowce chemiczne – jest kontrolowany przez garstkę wyspecjalizowanych firm. Wystarczy odciąć jedno ogniwo, by cała linia produkcyjna runęła. Dlatego sankcje technologiczne są tak chirurgicznie skuteczne. Zablokowanie sprzedaży maszyn holenderskiej ASML lub licencji amerykańskiej Cadence paraliżuje całe państwa.

Stany Zjednoczone zajmują tu unikalną pozycję, kontrolując nerw wojny cyfrowej: oprogramowanie EDA (Electronic Design Automation), czyli cyfrowe deski kreślarskie, na których powstają projekty układów (Synopsys, Cadence, Siemens EDA), oraz kluczowe architektury CPU/GPU (x86 Intela i AMD, oraz licencje ARM podlegające zachodnim regulacjom). To daje Waszyngtonowi potężne narzędzie w postaci tzw. **extraterritorial reach**, prawa, pozwalającego egzekwować swoje regulacje poza granicami USA. Nawet jeśli chip jest fizycznie produkowany w fabryce TSMC na Tajwanie, jego sprzedaż do Chin może zostać zablokowana, o ile do jego zaprojektowania użyto amerykańskiego oprogramowania.

Podręcznikowym przykładem tej strategii stał się los Huawei. Gdy w 2019 roku USA nałożyły sankcje, chiński gigant w jednej chwili stracił dostęp do najnowocześniejszych procesorów, które projektowała jego własna spółka HiSilicon, a produkował je tajwański TSMC. Efekt był natychmiastowy i druzgocący: gwałtowne wstrzymanie globalnej ekspansji sieci 5G i zapaść udziałów w rynku smartfonów, na którym Huawei był liderem.

Pekin odpowiedział strategią długiego marszu technologicznego. Z jednej strony, uruchomił gigantyczne inwestycje w krajowego czempiona, SMIC (Semiconductor Manufacturing International Corporation), oraz

dziesiątki mniejszych firm projektowych. Z drugiej strony, zaczął budować alternatywny ekosystem w oparciu o RISC-V, czyli otwartą, niepodlegającą amerykańskim licencjom architekturę procesorów. Równolegle Chiny pompują miliardy w rozwój własnego sprzętu do litografii DUV, próbując w kilka lat nadrobić dekady zapóźnienia. Rezultaty są jednak pełne sprzeczności. SMIC zaimponował światu, produkując w 2023 roku chip w procesie 7 nm bez dostępu do maszyn EUV, jednak jego wydajność i koszty są nieporównywalne z TSMC. Jednocześnie Huawei zaskoczył premierą telefonu Mate 60 Pro z własnym chipem Kirin 9000S, dowodząc, że sankcje spowalniają innowację, ale jej nie zatrzymują.

Przypadek Rosji to już nie spowolnienie, lecz technologiczna zapaść. Pozbawiona własnego przemysłu półprzewodnikowego i odcięta od zachodnich dostaw od 2022 roku, stała się studium totalnej izolacji. Dokumentacja zdobyta na ukraińskim froncie ujawniła ponurą prawdę: w zaawansowanych systemach uzbrojenia, takich jak rakiety Iskander, montowano chipy kanibalizowane z pralek i zmywarek. To nie anegdota, lecz twardy dowód, że bez dostępu do globalnego łańcucha dostaw Rosja nie jest w stanie samodzielnie podtrzymać nawet swojej technologii wojskowej.

W tym strategicznym potrzasku znalazły się Europa (ASML, Zeiss, STMicroelectronics) i Japonia (Tokyo Electron, Nikon, Canon). Choć w dużej mierze podporządkowały się amerykańskim restrykcjom, ich firmy tracą miliardowe kontrakty na największym rynku zbytu, jakim są Chiny. Ich obawa jest prosta i racjonalna: jeśli dziś odetniemy Chiny, jutro obudzimy się w świecie, w którym Chiny nas nie potrzebują, i my na zawsze stracimy kluczowego partnera.

Sankcje działają więc jak zaciągnięty hamulec ręczny: gwałtownie spowalniają, destabilizują, zmuszają do szukania objazdów, lecz nie gaszą silnika innowacji. Huawei i SMIC udowodniły, że obejścia są możliwe: przez wykorzystanie starszych technik, inżynieryjną kreatywność i bezdenne wsparcie państwa. Jednak sankcje skutecznie uniemożliwiają osiągnięcie masowej skali. Chiny mogą stworzyć jeden zaawansowany chip, lecz nie są w stanie produkować ich w milionach sztuk z wydajnością TSMC. Embargo nie jest przyciskiem „on/off” dla postępu, lecz nastawem zaworu na rurociągu uczenia się.

Geopolityka chipów: USA, Chiny, Europa, Tajwan

Embargo na półprzewodniki przestało być narzędziem doraźnym, a stało się stałym instrumentem globalnej polityki, przypominającym dawne blokady morskie. Stany Zjednoczone, wykorzystując swoją kontrolę nad oprogramowaniem do projektowania chipów (EDA), kluczowymi architekturami procesorów oraz monopolami w wąskich gardłach łańcucha dostaw (ASML, Nvidia), zbudowały ekosystem, w którym każdy przepływ technologii może zostać wstrzymany jedną decyzją polityczną. To broń precyzyjna, która nie niszczy całych gospodarek, lecz paraliżuje ich najbardziej newralgiczne sektory, takie jak sztuczna inteligencja, superkomputery czy sieci 5G.

Chińska odpowiedź na to wyzwanie to strategia autarkii długiego marszu, oparta na trzech filarach. Pekin wie, że nie wygra tej wojny w krótkim terminie, toteż inwestuje w rozwój własnych, choć mniej zaawansowanych litografii, budowę równoległego ekosystemu opartego na otwartej architekturze RISC-V oraz masowe wykorzystanie komponentów COTS, czyli gotowych, komercyjnych układów scalonych, w wojsku i kosmosie. To logika roju: zamiast jednego, wartego pół miliarda dolarów satelity o podwyższonej odporności, Chiny mogą wysłać setki małych CubeSatów na niską orbitę, wyposażonych w tańsze chipy. Nie każdy przetrwa, ale cała konstelacja będzie działać – to myślenie bliższe kulturze „ilość ma swoją jakość" niż zachodniej obsesji na punkcie perfekcji.

Europa znalazła się na strategicznym rozdrożu. Z jednej strony dysponuje monopolowymi aktywami, takimi jak holenderski ASML czy niemieckie firmy Zeiss i Trumpf, bez których nie ma nowoczesnej litografii. Z drugiej strony nie posiada własnych fabryk zdolnych do masowej produkcji w najwyższych węzłach technologicznych. Scenariusz optymistyczny zakłada, że dzięki inicjatywom takim jak EU Chips Act kontynent zbuduje własne moce produkcyjne, czego zwiastunem są inwestycje Intela w Magdeburgu i TSMC w Dreźnie. W scenariuszu pesymistycznym Europa zostanie zdegradowana do roli dostawcy narzędzi i chemikaliów, podczas gdy Azja i USA podzielą między siebie rynek faktycznej produkcji.

Najbardziej zapalny scenariusz dotyczy jednak Tajwanu, gdzie TSMC odpowiada za około 90% światowej produkcji najbardziej zaawansowanych chipów. Dla Stanów Zjednoczonych fabryki te stanowią „krzemową tarczę", gwarantującą, że Pekin nie zdecyduje się na inwazję, gdyż ich zniszczenie wywołałoby globalny kryzys gospodarczy. Dla Chin z kolei TSMC jest „krzemowym zakładnikiem". Wystarczy blokada morska wyspy lub paraliżujący cyberatak, by odciąć świat od blisko jednej trzeciej nowej mocy obliczeniowej rocznie.

Inni gracze pozostają daleko w tyle. Rosja jest już technologicznie wykluczona i bez szans na odbudowę własnego ekosystemu. Indie, mimo ogromnego potencjału w sektorze IT i oprogramowania, spóźniły się na pociąg z napisem „hardware". Izrael pozostaje graczem niszowym, silnym w projektach specjalistycznych. Tymczasem Korea Południowa i Japonia, choć są wiernymi sojusznikami USA, muszą uprawiać delikatny taniec, ponieważ ich giganci przemysłowi, tacy jak Samsung, Sony czy Nikon, są głęboko uzależnieni od chińskiego rynku zbytu.

Chiny: droga do chipowej samowystarczalności

Odpowiedzią Chin stał się długi marsz technologiczny, prowadzony na dwóch frontach. Z jednej strony to frontalny atak przez masowe inwestycje w krajowych czempionów, jak SMIC, i dziesiątki mniejszych firm projektowych. Z drugiej, to manewr oskrzydlający, czyli budowa alternatywnej ekosfery w oparciu o RISC-

V, otwartą architekturę procesorów. Pekin inwestuje również w produkcję sprzętu litograficznego DUV, próbując w dekadę nadrobić zaległości, które Zachód budował przez pół wieku.

Rezultaty tej kampanii pozostają jednak niejednoznaczne. SMIC w 2023 roku pokazał chip 7 nm wyprodukowany bez maszyn EUV – to imponujący wyczyn inżynierski, lecz jego wydajność i koszty są nieporównywalne z TSMC. Równocześnie Huawei zaskoczył świat premierą telefonu Mate 60 Pro z chipem Kirin 9000S. To dowód, że embargo nie jest guzikiem „on/off" dla postępu, tylko nastawem zaworu na rurociągu uczenia się; spowalnia, ale nie zatrzymuje innowacji.

Dżule na decyzję: nowa miara ekonomii chipów

Pojęcie całkowitego kosztu posiadania (TCO) w świecie półprzewodników przeszło cichą rewolucję. Dziś nie jest to już wyłącznie cena wafla krzemowego czy pojedynczego układu, lecz złożona suma nakładów energetycznych, kosztów serwisu, logistyki pakowania i, coraz częściej, politycznych kosztów licencji. Najprostszym sposobem pomiaru TCO nie są już dolary za chip, lecz dżule na przetworzony bit i sekundy dzielące do podjęcia decyzji.

W centrach danych sektora obronnego zużycie energii stało się fizyczną barierą rozwoju. Pojedynczy akcelerator klasy H100, wykonany w technologii 5 nm i wyposażony w pamięć HBM3, pobiera 700 watów; w klastrze liczącym dziesiątki tysięcy takich kart całkowite zapotrzebowanie rośnie do setek megawatów. W tym ujęciu ekonomia mocy obliczeniowej zaczęła przypominać energetykę szczytową, gdzie rezerwuje się sloty serwerowe niczym godziny szczytowego poboru w elektrowni. Cena decyzji to nie tylko koszt sprzętu, ale także chłodzenia, redundancji i utrzymania.

Na orbicie ten rachunek staje się jeszcze bardziej bezwzględny, gdyż każdy dżul energii trzeba dostarczyć z paneli słonecznych lub baterii. Dlatego kluczową miarą wydajności stają się „operacje na dżul", a układy typu rad-hard, czyli odporne na promieniowanie, wybiera się nie ze względu na cenę, lecz zdolność do unikania restartu całej misji. Koszt pojedynczego błędu bitowego to nie utrata pliku, lecz strata satelity wartego setki milionów dolarów. Logika jest więc prosta: lepiej zapłacić dziesięć razy więcej za chip, jeśli minimalizuje on ryzyko utraty całej platformy.

W domenie wojskowej TCO objawia się w koszcie decyzji taktycznej. Bomba Paveway w Wietnamie kosztowała znacznie więcej niż jej „głupi" odpowiednik, ale drastycznie obniżała liczbę lotów bojowych i związane z nimi ryzyko polityczne. Pocisk Tomahawk był drogi w przeliczeniu na sztukę, ale tani w skali całej kampanii, ponieważ ograniczał konieczność wysyłania załogowych samolotów. Dziś akcelerator GPU z pamięcią HBM to cyfrowy odpowiednik Paveway: jednostkowo drogi, ale tani w przeliczeniu na misję, bo skraca pętlę decyzyjną i zmniejsza ryzyko katastrofalnego błędu.

Paradoks chipów: tanie tranzystory, drogie fabryki

Pierwsze układy scalone Jacka Kilby'ego i Roberta Noyce'a kosztowały jak kruszec. Texas Instruments oferowało je po 1000 dolarów za 64 sztuki, co dawało astronomiczną kwotę 15 dolarów za pojedynczy chip. Fairchild próbował podobnej strategii, dopóki Noyce nie wykonał ruchu, który w każdej szkole biznesu uznano by za samobójczy: obniżył ceny do 2 dolarów, często nurkując poniżej kosztów produkcji. Paradoksalnie, to rynkowy kanibalizm – sprzedaż ze stratą – zbudował fundament pod cywilną rewolucję. Układ scalony przestał być zabawką dla wojska, a stał się krwiobiegiem gospodarki, od kalkulatorów po telewizory.

Kluczem do tej strategii była krzywa uczenia się, czyli zasada, zgodnie z którą każde podwojenie wolumenu produkcji obniża koszty o kilkanaście procent. Andy Grove w Intelu narzucił żelazną regułę „dokładnego kopiowania": jeśli jedna linia produkcyjna osiągała lepsze wyniki, jej procesy replikowano co do joty we wszystkich fabrykach. Morris Chang w Texas Instruments traktował produkcję jak problem statystyczny, gdzie celem była minimalizacja odchyleń i odrzutów. W Micronie w latach 80. obsesyjnie zmniejszano chipy i modyfikowano piece, by wypalać więcej wafli naraz. Każdy trik, każde pół procenta zysku na wydajności, stawało się kwestią przetrwania.

Rezultat tej presji był sejsmiczny. Cena pojedynczego tranzystora spadła w ciągu półwiecza o miliard razy, co stanowi spadek bardziej dramatyczny niż w energetyce czy transporcie. Gdyby lotnictwo cywilne rozwijało się w tym tempie, bilet z Warszawy do Tokio kosztowałby dziś mniej niż zapałki, a transkontynentalny lot zużywałby tyle paliwa, co zapalniczka.

W ten sposób narodził się fundamentalny paradoks tej branży: pojedynczy tranzystor jest niemal darmowy, ale fabryka zdolna go wyprodukować kosztuje fortunę rosnącą w tempie wykładniczym. Zakład Intela czy TSMC na węźle 5 nm to inwestycja rzędu 20 miliardów dolarów, a pojedyncza maszyna EUV od ASML kosztuje ponad 150 milionów. Jej następna generacja, High-NA EUV, ma być wyceniana na 350 milionów za sztukę. To podręcznikowy przykład monopolu naturalnego, czyli sytuacji, w której gigantyczne koszty wejścia i ekonomia skali nieuchronnie prowadzą do koncentracji rynku. W efekcie na szczycie globalnej piramidy technologicznej zostało miejsce tylko dla trzech graczy: TSMC, Samsunga i Intela.

Moc obliczeniowa: od ENIAC-a do potęgi państw

W 1945 roku armia amerykańska uruchomiła ENIAC, pierwszy cyfrowy komputer wielkiej skali. Kolos ważący 30 ton wypełniał po brzegi halę fabryczną i pożerał 160 kilowatów energii, a jego moc obliczeniowa, mierzona w tysiącach operacji na sekundę, wystarczała zaledwie do kalkulacji trajektorii pocisków

artyleryjskich. Już wtedy jednak zrozumiano, że liczby mogą zmieniać bieg wojen. ENIAC był narzędziem wojskowym, nie cywilnym, a cała informatyka odziedziczyła po nim ten militarny rodowód.

Wynalezienie tranzystora, a następnie układu scalonego, fundamentalnie zmieniło reguły gry. Gordon Moore w 1965 roku sformułował obserwację, że liczba elementów w układach scalonych będzie się podwajać co dwa lata. To proroctwo, ochrzczone później Prawem Moore'a, stało się najtrafniejszą prognozą technologiczną XX wieku. Dzięki niemu moc obliczeniowa rosła wykładniczo, a koszt pojedynczej operacji spadał w tempie, które trudno objąć wyobraźnią – o miliardy procent w ciągu kilku dekad.

Pierwsze mikroprocesory, jak Intel 4004 z 1971 roku, były uniwersalnymi „mózgami" komputera, czyli CPU (Central Processing Unit). Jednak rosnące zapotrzebowanie wymusiło specjalizację. Tak narodziły się GPU (Graphics Processing Unit), układy zaprojektowane do renderowania grafiki, które okazały się fenomenalnie skuteczne w obliczeniach równoległych. CPU przypomina genialnego profesora rozwiązującego jedno złożone zadanie, podczas gdy GPU to armia rekrutów wykonująca jednocześnie tysiące prostych operacji. Nvidia, tworząc swoje procesory, nie mogła przewidzieć, że kładzie fundament pod rewolucję AI, która opiera się właśnie na masowym przetwarzaniu równoległym.

Dziś to GPU są sercem sztucznej inteligencji. Modele uczenia maszynowego wymagają bilionów operacji na macierzach, a architektura GPU wykonuje je tysiące razy szybciej niż tradycyjne CPU. Dlatego to nie Intel, weteran ery PC, lecz Nvidia, specjalista od grafiki, stała się technologicznym symbolem XXI wieku.

W latach 90. i dwutysięcznych potęgę państw zaczęto mierzyć w superkomputerach. Już w latach 70. amerykański Illiac IV analizował dane sonarowe, zwiększając wykrywalność radzieckich okrętów podwodnych. Później nastała era wyścigu o petaflopy, czyli biliardy operacji na sekundę, a dziś rywalizacja toczy się już w skali exaflopów – trylionów operacji. Chiński Sunway TaihuLight czy amerykański Frontier to nie tylko narzędzia naukowe, ale i instrumenty siły militarnej. Służą do symulacji eksplozji jądrowych, projektowania broni hipersonicznej i analizy danych wywiadowczych. Każdy nowy rekord mocy obliczeniowej to geopolityczny komunikat: nasze algorytmy są szybsze od waszych.

Jednak siła nie tkwi już tylko w scentralizowanych superkomputerach. Chmura obliczeniowa stała się nową formą suwerenności. Amazon AWS, Microsoft Azure i Google Cloud to cyfrowe supermocarstwa, dysponujące centrami danych o potędze przekraczającej zasoby wielu państw. Wynajmują moc obliczeniową w modelu usługowym, ale są też partnerami strategicznymi rządów. Pentagon, poprzez kontrakty takie jak JEDI czy JWCC, deleguje część swojej analitycznej siły do prywatnych, lecz kluczowych serwerowni.

Dziś symbolem potęgi nie jest dywizja pancerna, ale klaster kart graficznych H100 Nvidii. Każda kosztuje dziesiątki tysięcy dolarów, a zamówienia Pentagonu czy chińskich gigantów technologicznych idą w tysiące sztuk. Stany Zjednoczone, świadome tej zależności, zakazały eksportu najnowszych GPU do Chin, próbując odciąć rywalowi tlen napędzający rozwój AI i symulacji militarnych. Chińskie firmy, jak Huawei, próbują

tworzyć alternatywy, ale pozostają w tyle. Arsenał XXI wieku nie jest liczony w głowicach, lecz w dostępie do mocy obliczeniowej.

Tymczasem Prawo Moore'a dobiega kresu. Miniaturyzacja tranzystorów do skali kilku nanometrów dociera do granic fizyki kwantowej i staje się ekonomicznie absurdalna. Inżynierowie szukają więc nowych paradygmatów: układają tranzystory w trójwymiarowe wieżowce (FinFET, gate-all-around) lub eksperymentują z zupełnie nowymi modelami, jak komputery neuromorficzne czy kwantowe. Jak ujął to Jim Keller, legendarny projektant z AMD i Intela: „Możliwości rozwoju jeszcze nie zostały wyczerpane. Granice są w naszej wyobraźni, nie w krzemie".

W 1945 roku siłę mierzono w tonach stali. W epoce zimnej wojny – w megatonach. Dziś mierzy się ją w exaflopach i terabajtach na sekundę. Państwo z przewagą w mocy obliczeniowej zyskuje dominację w sztucznej inteligencji, projektowaniu broni, analizie big data i eksploracji kosmosu. Moc obliczeniowa stała się niewidzialną grawitacją geopolityki. To ona definiuje, kto krąży w centrum globalnego systemu, a kto zostaje zepchnięty na jego peryferie.

Niezawodność chipów: od rad-hard do zero trust

Pierwsze chipy były równie kapryśne jak lampy próżniowe, które miały zastąpić. W Texas Instruments inżynierka Mary Anne Potter spędzała tygodnie, testując próbki tranzystorów i ręcznie prowadząc analizy regresji. Wczesne układy cierpiały na wrodzoną ułomność: zanieczyszczenia na powierzchni krzemu wchodziły w reakcje z materiałami, prowadząc do zwarć i awarii. Przełomem okazała się metoda planarna Jeana Hoerniego, w której tranzystory powstawały w całości wewnątrz krzemowego podłoża, chronione hermetyczną warstwą tlenku krzemu. To rozwiązanie poprawiło niezawodność o rząd wielkości i otworzyło drogę do pierwszych masowych kontraktów wojskowych.

W latach 60. montaż i testowanie chipów przypominały pracę w zakładzie jubilerskim. Krzemowy rdzeń łączono z zewnętrznymi wyprowadzeniami za pomocą cienkich złotych drucików, a każdy gotowy układ podłączano ręcznie do aparatury pomiarowej. Zadania te, wymagające niezwykłej precyzji i cierpliwości, często powierzano kobietom. Automatyzacja nie istniała, toteż każda partia wafli krzemowych stanowiła mały, niepowtarzalny eksperyment, którego wynik był daleki od pewności.

Dwie dekady później Japonia uczyniła z jakości swoją kulturową przewagę. W latach 80. producenci pamięci DRAM z Tokio osiągali wskaźniki awaryjności na poziomie 0,02% w ciągu pierwszego tysiąca godzin pracy. W tym samym czasie najlepsze firmy amerykańskie notowały 0,09%, a najgorsze nawet 0,26%. Nie była to różnica kosmetyczna, lecz fundamentalna przewaga kosztowa i reputacyjna. Amerykańskie wojsko, dla

którego niezawodność była dogmatem, zaczęło bić na alarm. Pentagon wsparł programy jakościowe, a konsorcjum Sematech powstało właśnie jako strategiczna odpowiedź na japońską perfekcję.

Wraz z globalizacją na scenę wkroczyły firmy OSAT, czyli Outsourced Semiconductor Assembly and Test, które przejęły masowe testowanie i pakowanie układów. Weryfikacja nie kończyła się bowiem na poziomie wafla krzemowego; każdy gotowy chip musiał przejść rygorystyczne procedury, zanim trafił do smartfona czy satelity. Azja Południowo-Wschodnia – z Tajlandią, Malezją i Filipinami na czele – stała się globalnym centrum tej działalności, łącząc relatywnie tanią siłę roboczą z postępującą automatyzacją.

W epoce litografii EUV, gdy pojedyncza maszyna ASML kosztuje ponad 150 milionów dolarów, testowanie stało się równie krytyczne jak sama produkcja. Każda godzina przestoju generuje straty liczone w tysiącach dolarów, toteż ASML wdrożyło politykę „ekstremalnej niezawodności”. Każdy komponent musi wytrzymać średnio 30 000 godzin pracy bez awarii. Do gry weszła konserwacja prognozowana, czyli algorytmy analizujące dane z czujników, by przewidzieć usterkę, zanim ona wystąpi. To logika przeniesiona wprost z lotnictwa do serca fabryki półprzewodników.

Zastosowania wojskowe wymagają jednak testów o zupełnie innej naturze. Od początku XXI wieku Pentagon obawia się hardware’owych backdoorów – ukrytych luk wprowadzanych celowo na etapie produkcji lub montażu. Agencja DARPA finansuje programy weryfikacyjne, których celem jest zagwarantowanie, że każdy chip jest wolny od wrogiej manipulacji. Tak narodziła się doktryna „zero trust”, czyli nie ufaj niczemu, weryfikuj wszystko. W praktyce oznacza to wielowarstwowe testy, od analizy topologii krzemu po wszczepianie mikroczujników monitorujących układ przez cały cykl jego życia. W wojsku nie chodzi już tylko o niezawodność techniczną, ale o bezpieczeństwo strategiczne.

W przestrzeni kosmicznej testy stają się brutalne. Chipy muszą przetrwać bombardowanie protonami, ciągłą ekspozycję na promieniowanie kosmiczne i ekstremalne cykle termiczne, od -150 do +150°C. Testy kwalifikacyjne trwają miesiącami i obejmują symulowane warunki radiacyjne. Stąd fenomen układów typu rad-hard: drogich, powolnych, ale niemal nieśmiertelnych w próżni. W przypadku konstelacji na niskiej orbicie okołoziemskiej (LEO) przyjęto inną logikę. Zamiast testować każdy chip na dekady, stosuje się redundancję z użyciem tanich komponentów COTS, które można łatwo zastąpić przy kolejnych wystrzeleniach.

Testowanie chipów nie jest nudną kontrolą jakości, lecz integralną częścią ich historii. Ewolucja prowadzi od Mary Anne Potter, która ręcznie „debugowała” tranzystory, po algorytmy predykcyjne w maszynach EUV. To przejście od rzemiosła i intuicji do matematyki i sztucznej inteligencji. W tle tej historii zawsze stała polityka. W zimnej wojnie stawką była niezawodność pocisku Minuteman, dziś chodzi o pewność, że w krzemie nie ukryto cyfrowego szpiega.

Ewolucja prowadzi od Mary Anne Potter, która ręcznie „debugowała” tranzystory, po algorytmy predykcyjne w maszynach EUV. To przejście od rzemiosła i intuicji do matematyki i sztucznej inteligencji. W tle tej historii zawsze stała polityka. W zimnej wojnie stawką była niezawodność pocisku Minuteman, dziś chodzi o pewność, że w krzemie nie ukryto cyfrowego szpiega.

Inspiracje

[1] Mary Anne Potter

APA ONE Publishing System

Article ID: 0b3ab988-8f39-4f76-a25c-780021e421ca