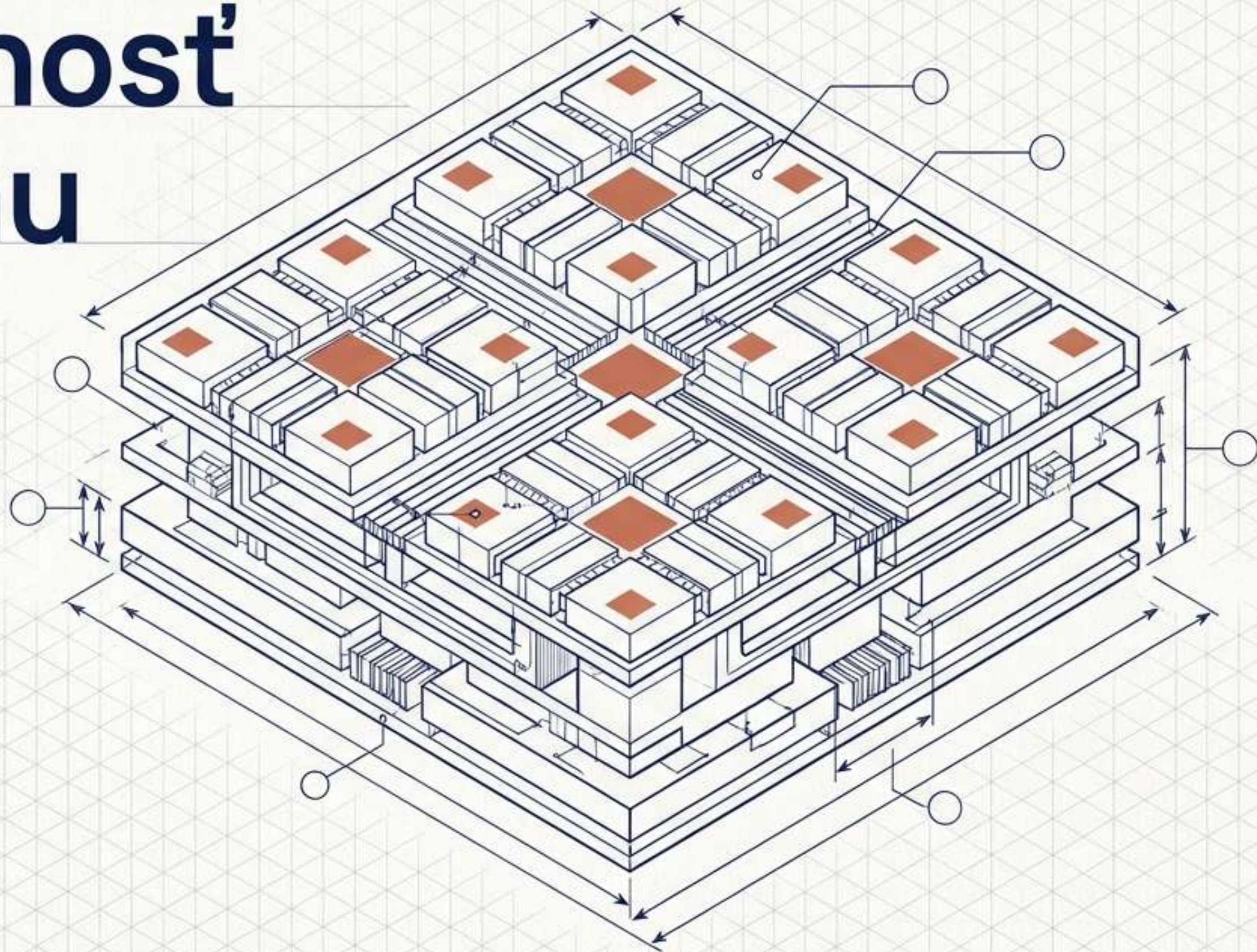


Nevyhnutnosť Paralelizmu

Prečo fyzika a softvérový
dopyt navždy zmenili
architektúru počítačov.



Na základe prednášok:
Paralelné počítačové systémy
(Norbert Ádám)

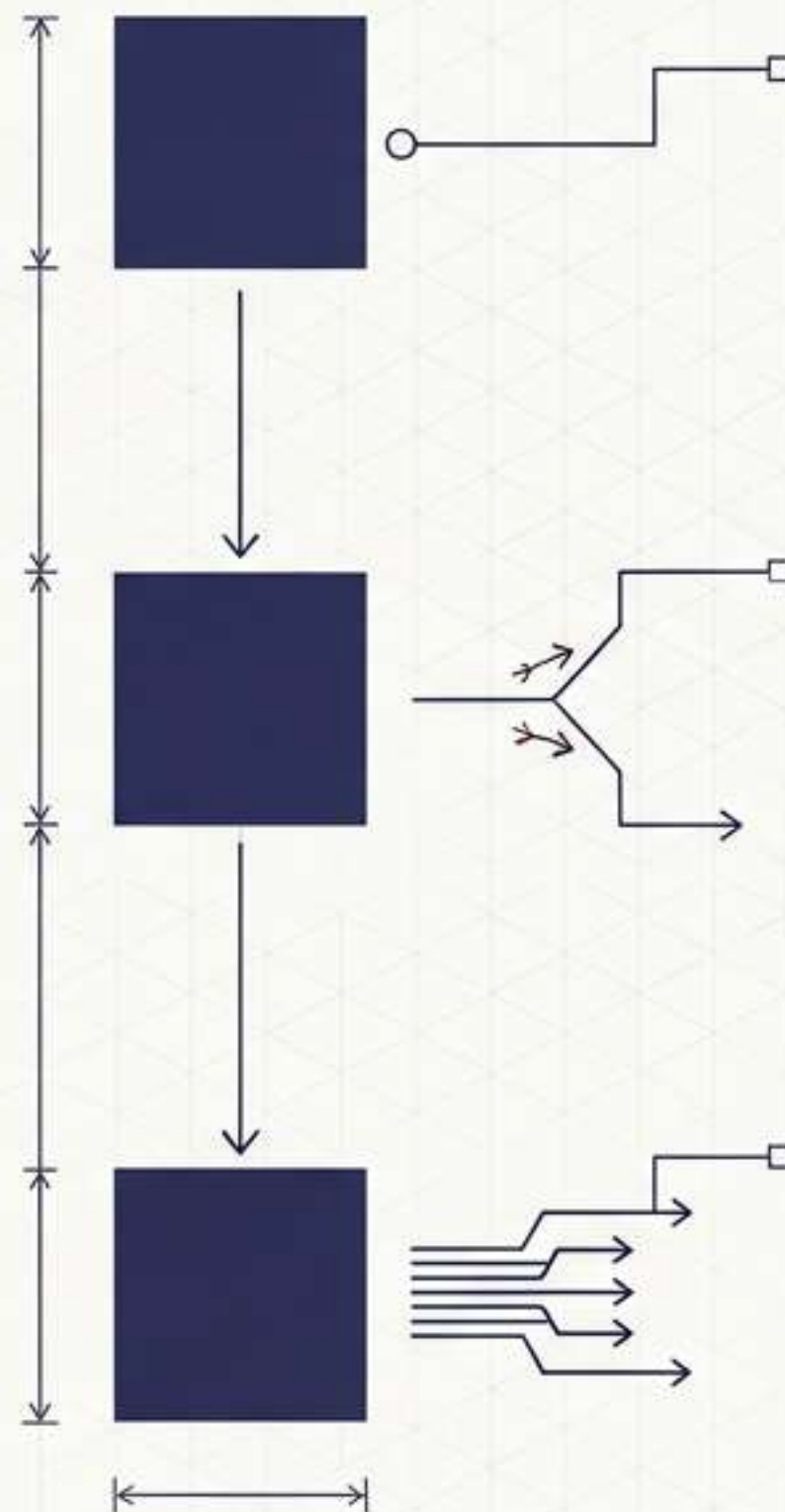
Koniec éry frekvencie

Zvyšovanie taktu
procesorov narazilo
na svoje

fyzikálne limity.

Ďalší rast výkonu
si vyžaduje

**zásadnú zmenu
architektonickej
perspektívy.**



Jedna technológia:

Notebooky a superpočítače sú
dnes z princípu veľmi podobné.

Konvergencia:

Technologické trendy
spôsobujú zjednocovanie
architektonických prístupov.

Nová nutnosť:

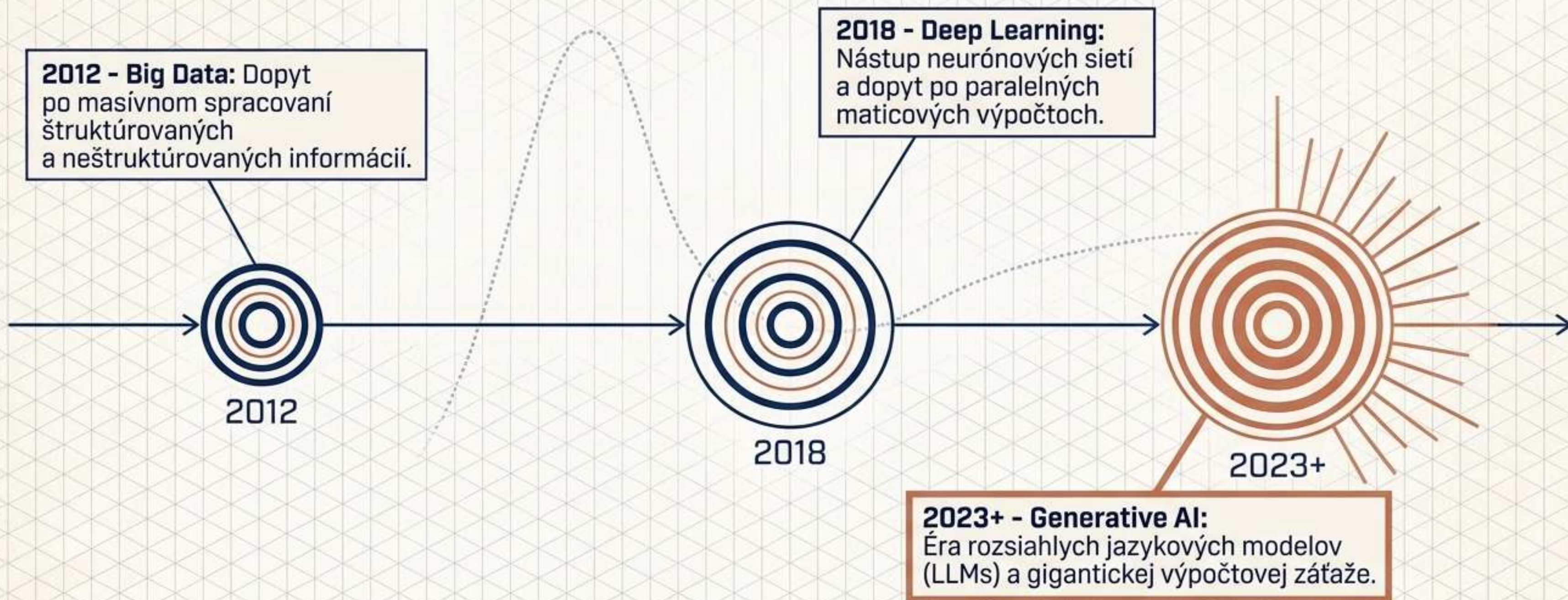
Paralelné počítanie už nie je
špecializovanou doménou,
ale absolútnou nevyhnutnosťou.

Anatómia technologického dopytu

Ako Gartner Hype Cycle mapuje životný cyklus inovácií



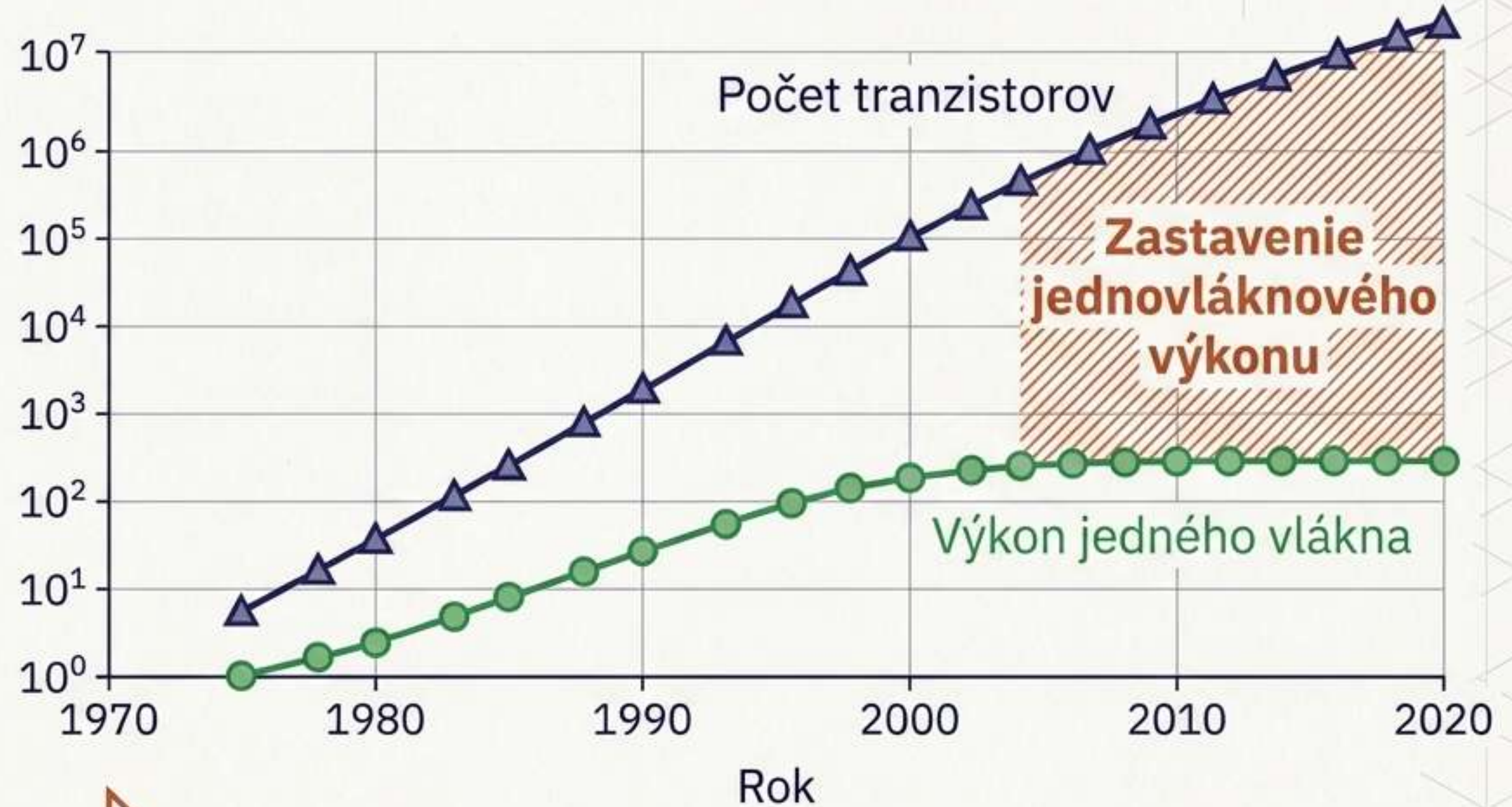
Eskalácia softvérového dopytu (2012 – 2025)



Od simulácií CFD a CAD až po umelú inteligenciu – moderné aplikácie vyžadujú priepustnosť, ktorú jednovláknový výkon nedokáže poskytnúť.

Hardvérový paradox

- Výkon mikroprocesorov historicky rástol 1,2-násobne ročne.
- Počet tranzistorov sa zdvojnásobuje každých ~18 mesiacov (Mooreov zákon).
- **Míľnik:** V roku 2015 úroveň výpočtového výkonu dosiahla ekvivalent mozgu myši.

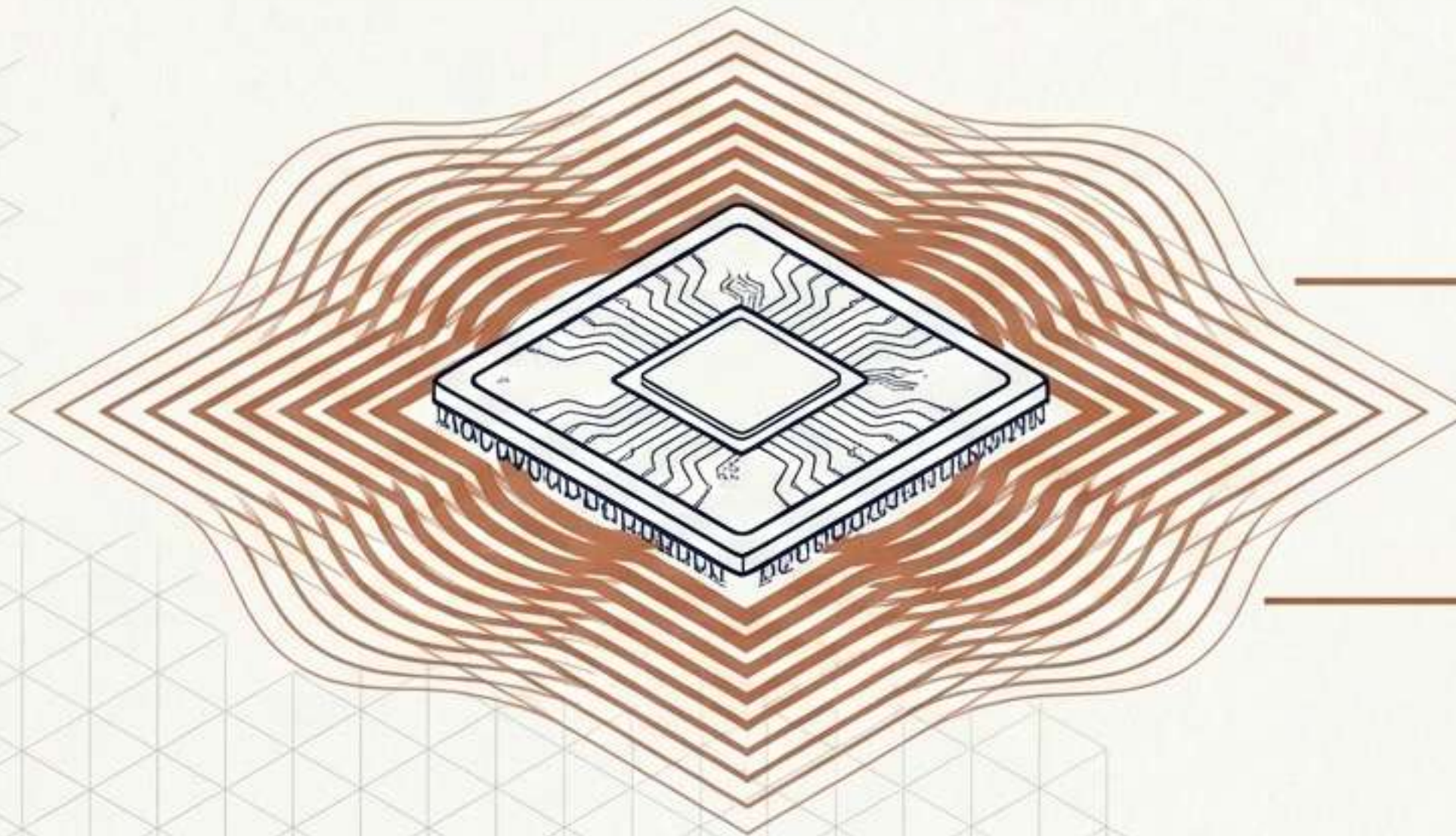


Otázka architekta: Ako efektívne využiť rastúci počet tranzistorov, keď ich nemôžeme taktovať rýchlejšie?

Fyzikálny náraz #1: Tepelná stena (Power Wall)

$$P \sim C_{\text{dyn}} \cdot V^2 \cdot f$$

Rovnica spotreby ukazuje, že spotreba (P) rastie priamo úmerne s frekvenciou (f), ale s kvadrátom napätia (V).



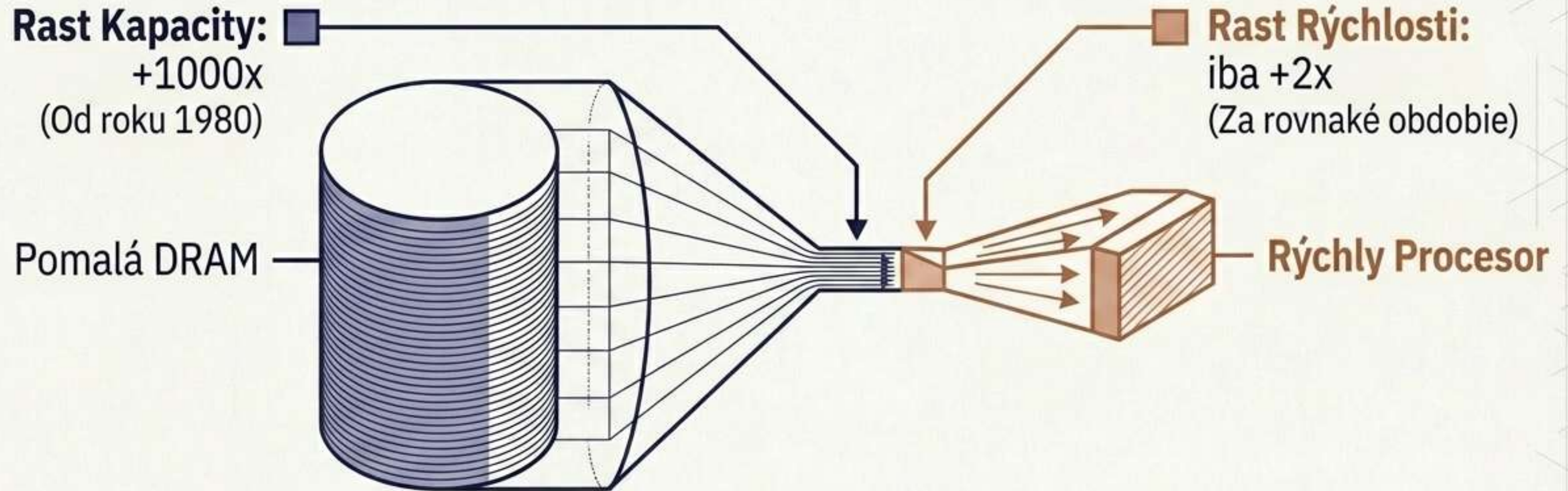
Fyzický limit: Čip o veľkosti $1,5 \times 1,5$ cm generuje vyše 130 W tepla.

Bariéra: Toto predstavuje absolútnu hranicu chladenia bežným vzduchom.

Dôsledok: Ak by sme ďalej zvyšovali frekvenciu, čip by sa fyzicky roztavil.

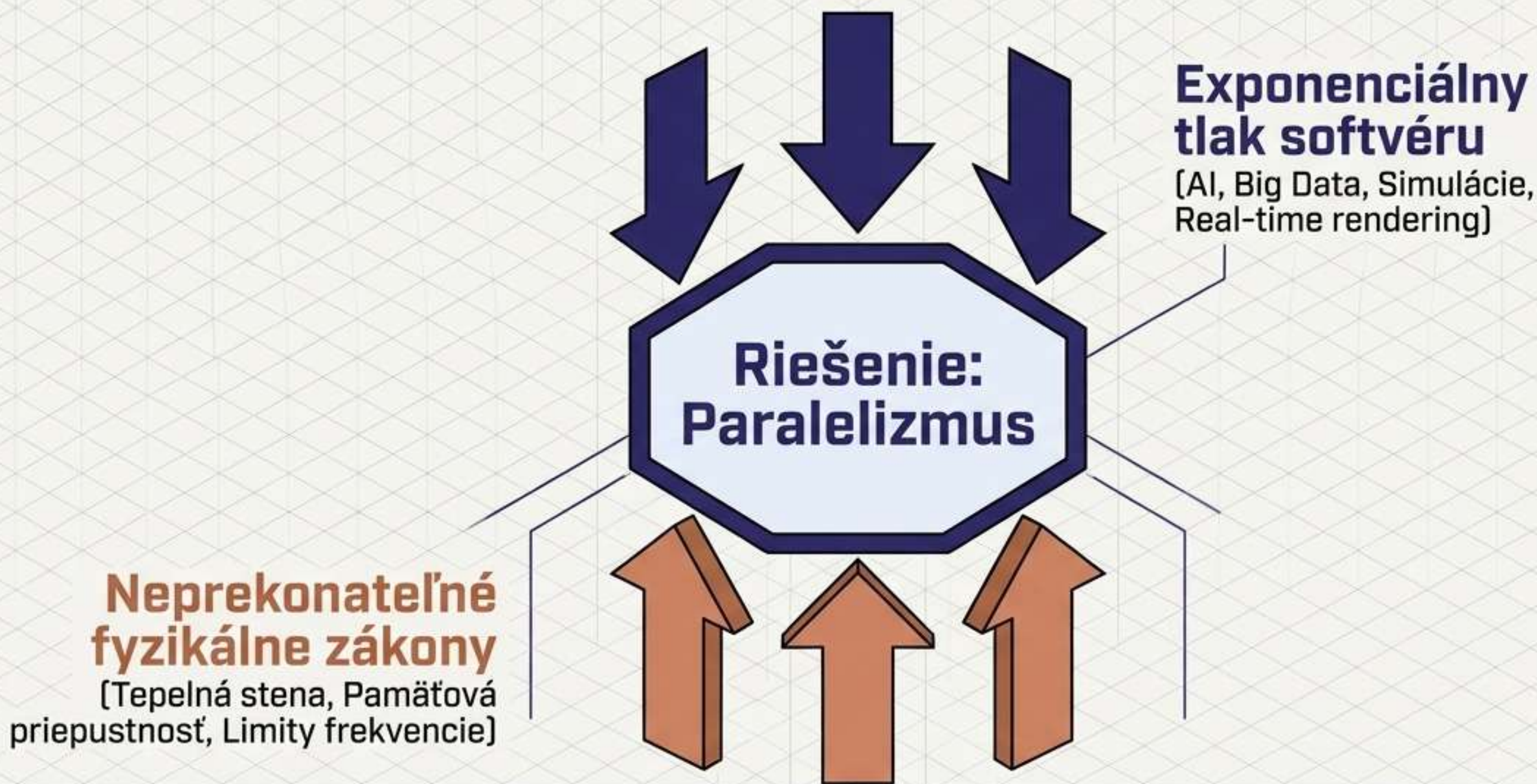
Fyzikálny náraz #2: Pamäťová stena (Memory Wall)

Extrémne rýchly procesor čakajúci na pomalé dáta.



Otvárajúce sa nožnice. Väčšie pamäte sú inherentne pomalšie. Aby sme nakrmili rýchle procesory, musíme začať prenášať dáta paralelne cez širšie rozhrania a hlbšie hierarchie (Cache).

Zrážka dopytu s fyzikou



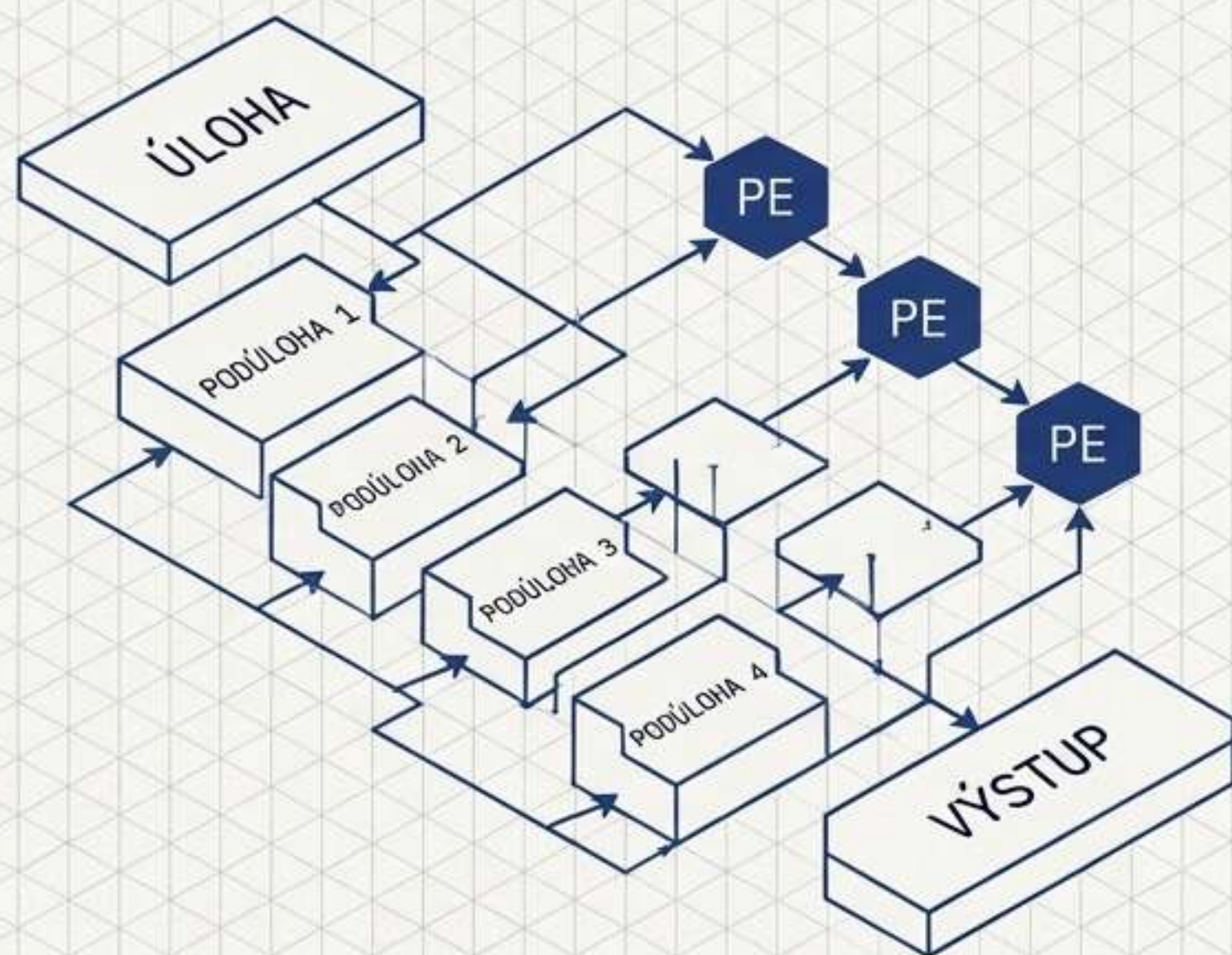
Paralelizmus nie je len jedným z možných inžinierskych prístupov. V tomto „tlakovom hrnci“ sa stal jedinou fyzikálne a ekonomicky životaschopnou cestou vpred.

Zmena paradigmy

Stará paradigma (Single-Thread)	Nová paradigma (Multi-Core / Paralelizmus)
Zvyšovanie taktovacej frekvencie (GHz)	Zvyšovanie celkovej priepustnosti systému
Zložitejšia logika pre zrýchlenie jednej inštrukcie	Replikácia výpočtových jadier (Hrubozrnný paralelizmus)
Hardvér zrýchli kód automaticky (zadarmo)	Nutnosť explicitne písať viacvláknový (paralelný) kód
Rýchly náraz na tepelnú stenu (Roztavenie)	Náročnosť na synchronizáciu a distribúciu zdrojov

Čo je paralelná architektúra?

Paralelný počítač je súbor procesných elementov (PE), ktoré aktívne spolupracujú pri efektívnom riešení masívnych výpočtových problémov.



Alternatíva k frekvencii:

Je to hlavný nástroj na zvyšovanie výkonu bez nutnosti zvyšovať takt.

Všadeprítomnosť:

Uplatňuje sa na všetkých úrovniach návrhu (od inštrukcií cez vlákna až po celé servery).

Nová perspektíva:

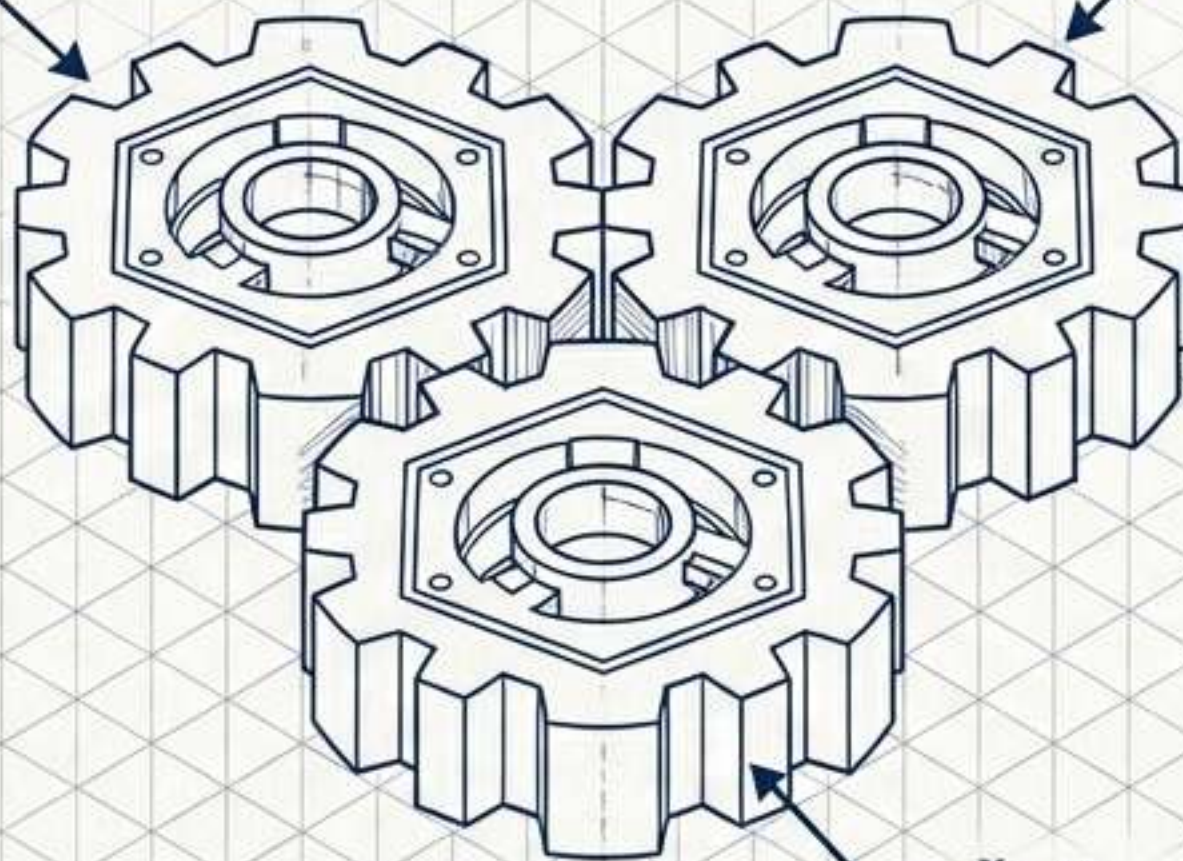
Je to smerodajný uhol pohľadu, z ktorého dnes architekti nazerajú na akýkoľvek výpočtový systém.

Anatómia paralelného systému

Úloha architekta: Maximalizovať výkon a programovateľnosť v rámci technologických limitov.

Alokácia zdrojov

- Aká veľká má byť architektúra?
- Aký výkonný má byť každý prvok?
- Koľko pamäte je potrebné prideliť?



Komunikácia a Synchronizácia

- Ako prvky navzájom spolupracujú?
- Ako sa prenášajú dáta medzi procesormi?

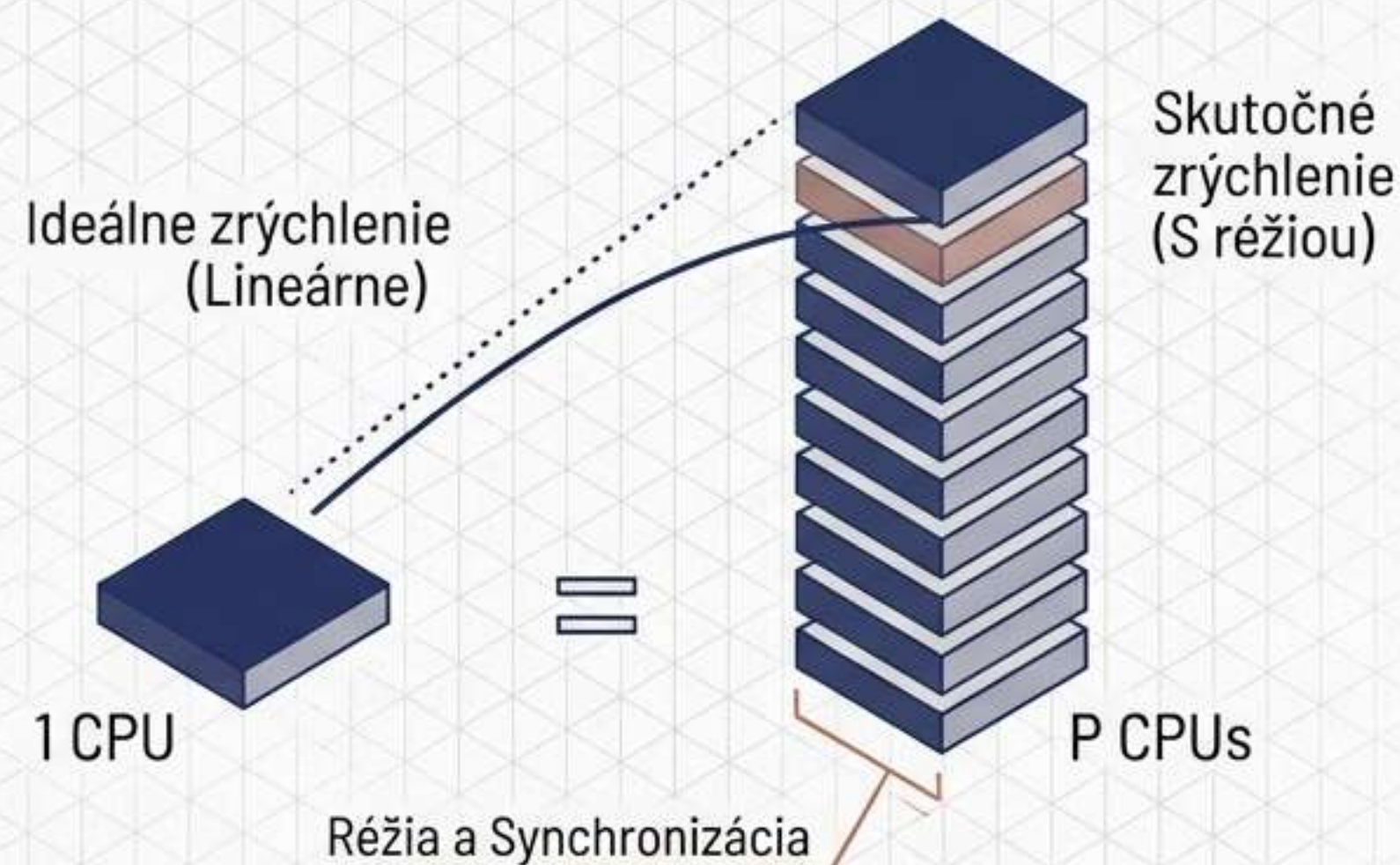
Škálovateľnosť

- Ako sa tieto rozhodnutia premietnu do reálneho výkonu?
- Dokáže systém efektívne rásť s veľkosťou problému?

Ultimátna metrika: Zrýchlenie (Speedup)

Cieľom aplikácií pri použití paralelných strojov je jediné – **zrýchlenie**. Pre fixnú veľkosť problému platí, že **výkon = 1 / čas**.

$$\text{Speedup}(p) = \frac{\text{Výkon s } (p) \text{ procesormi}}{\text{Výkon s } (1) \text{ procesorom}}$$



Dokonalý paralelizmus by znamenal, že p procesorov vyrieši úlohu p -krát rýchlejšie. Úlohou hardvérových abstrakcií a synchronizačných primitív je minimalizovať straty (réžiu), aby sme sa k tomuto ideálu čo najviac priblížili.

Nový štandard výpočtovej techniky

Jednotný princíp:

Od bežných desktopových systémov až po najväčšie superpočítačové klastre – všetky dnes stoja na viacjadrovom, paralelnom základe.

Rastúca priepustnosť:

Komerčné aplikácie (databázy, transakčné systémy, financie) už nevyžadujú len rýchlosť jedného úkonu, ale schopnosť spracovať tisíce úkonov naraz.

Nevyhnutná budúcnosť:

Dokiaľ nevznikne úplne nový fyzikálny model (napríklad kvantové počítanie), paralelizmus zostáva jediným plánom, ako nakrmiť neustále rastúci technologický hlad ľudstva.

