ECE7115 ~~Multimodal VLM~~ LLM

6. Scaling Laws

Spring 2026

Namhyuk Ahn, Inha University

Scaling

- Imagine the following scenario…
  - *Your friend has given you 10,000 H100s for a month, and asked you to build a good open source LLM*

- What do you do?
  - Put together your infra team and distributed training framework
  - Put together a great pre-training dataset
  - Run a big model (but which one??) ← we will learn this here

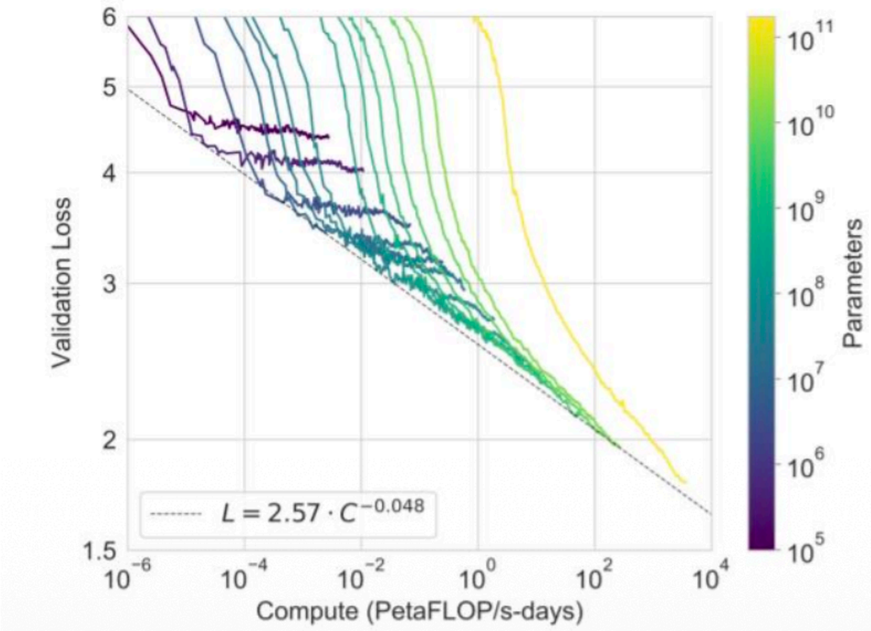
How to Scale Models?

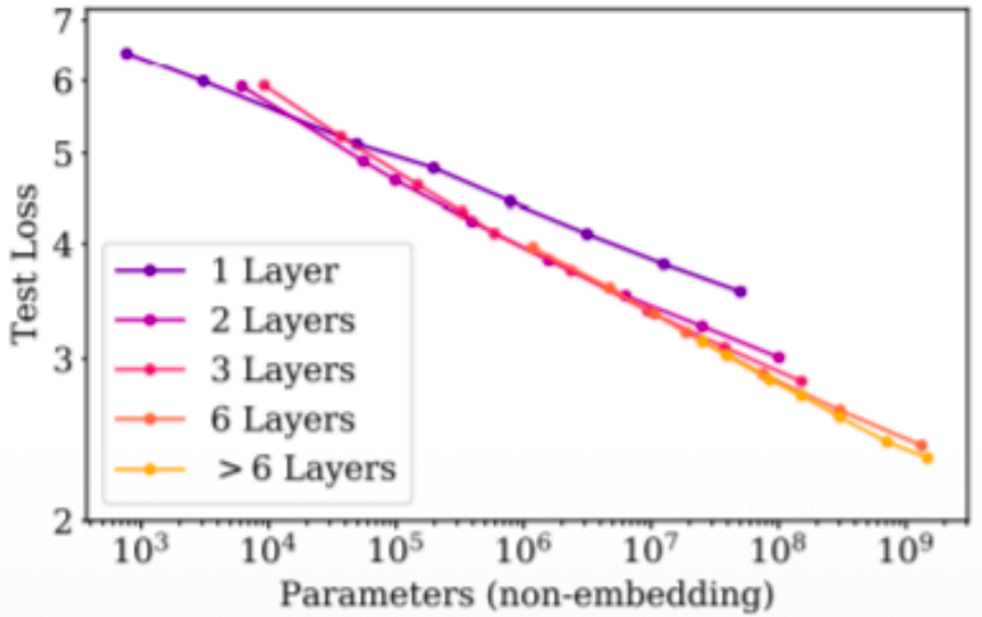
- We will probably use Transformer, but,
  - Wide or deep model?
  - Which model architecture?
  - How many heads?
  - Which nonlinearity?
  - How to search optimal hyperparameters?

- We could follow things from existing LLMs…
  - but how do these get optimized in the first place?

Today: Simple, Predictive "Laws"

- Scaling laws: Simple and predictive rules for model performance
  - Old way: tune hyperparameters on big models
  - New way: tune on small models, extrapolate to large ones





Scaling Laws

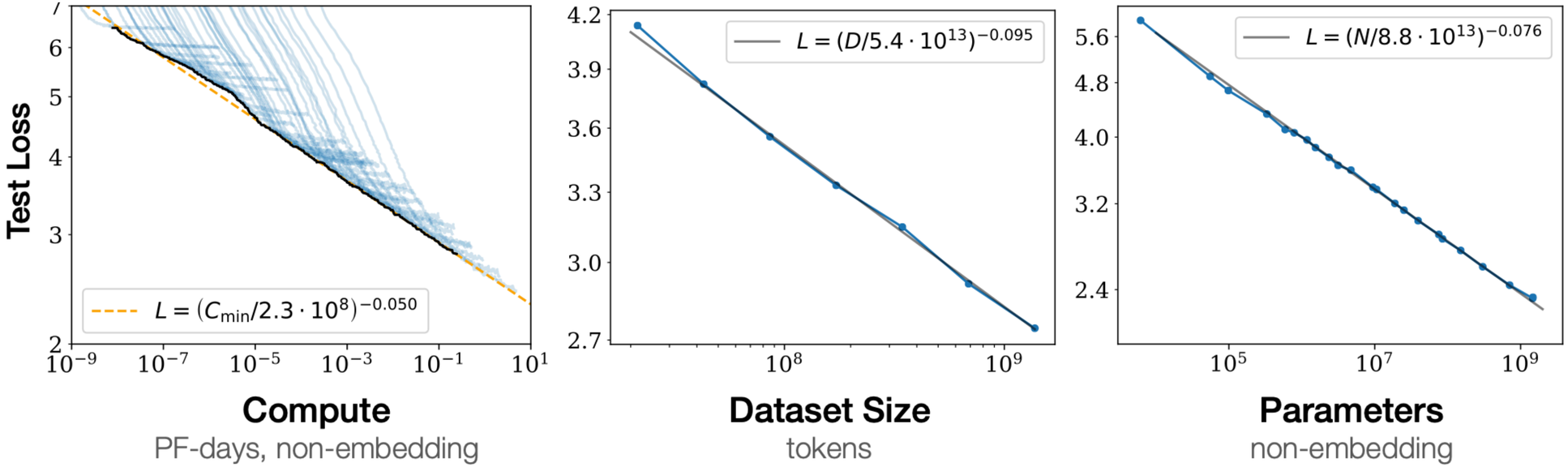
- Single Variable Scaling Laws
  - Are there rules that determine how data/model affects performance?
  - Data vs. Performance
  - Model size vs. Performance

- Scaling Laws for Model Engineering
  - How should we set hyperparameters on the big model?

- Compute Allocation
  - Do we train on more data or bigger models?
  - Data vs. Model size; which one is better?

Scaling Laws [Kaplan+ 2020]

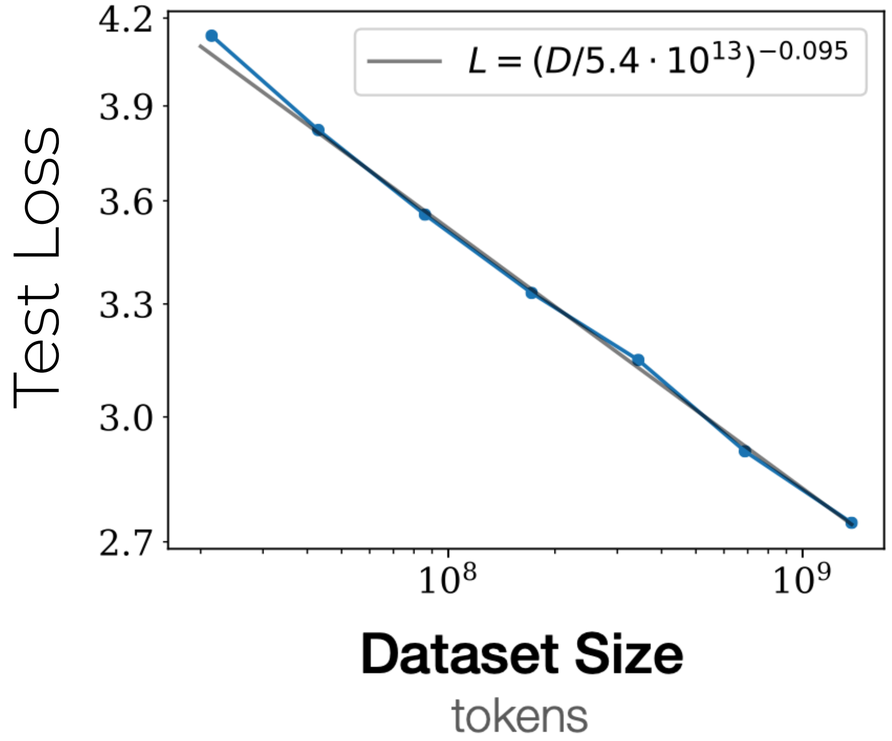
- Power law relationships for many factors
  - These scaling laws hold on many different kinds of phenomena!



Single Variable Scaling Laws

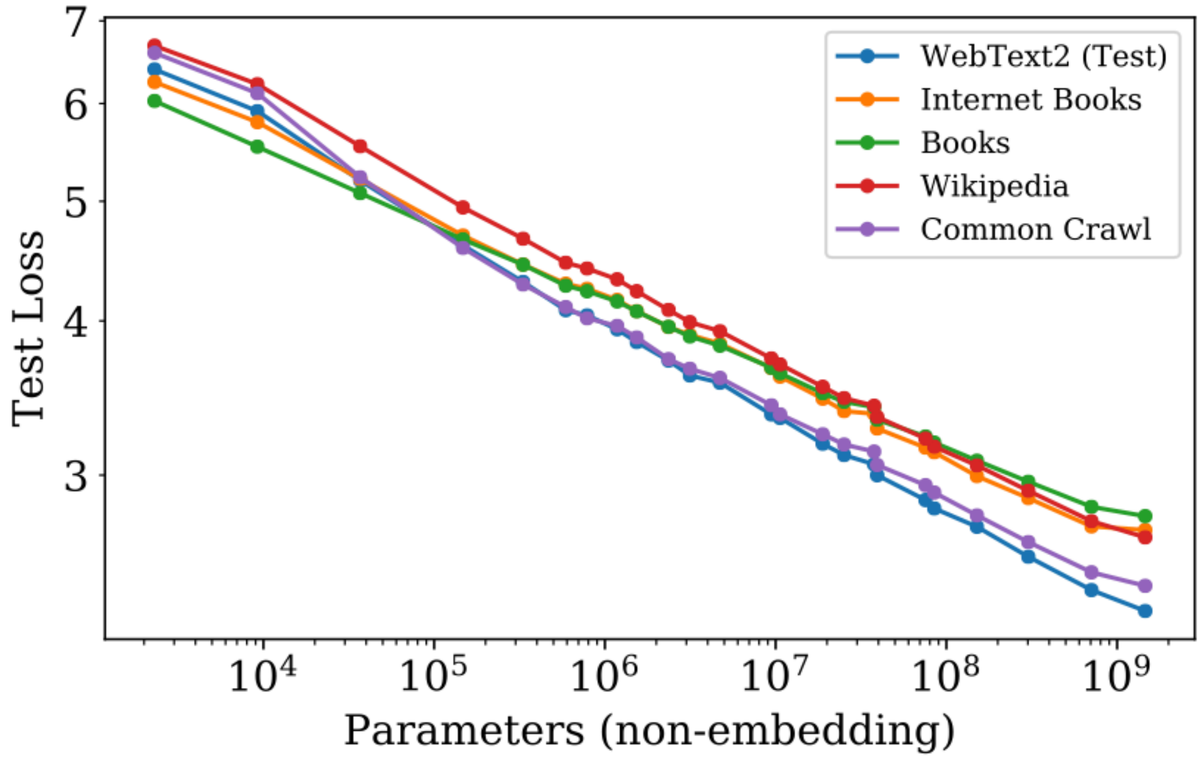
Data Scaling Laws [Kaplan+ 2020]

- TL;DR: Performance has a power-law relationship with data size
  - With a limited dataset with early stopping
  - $L(D) = (D_c/D)^{\alpha_D}$; $\alpha_D \sim 0.095, \quad D_c \sim 5.4 \times 10^{13}$ (tokens)



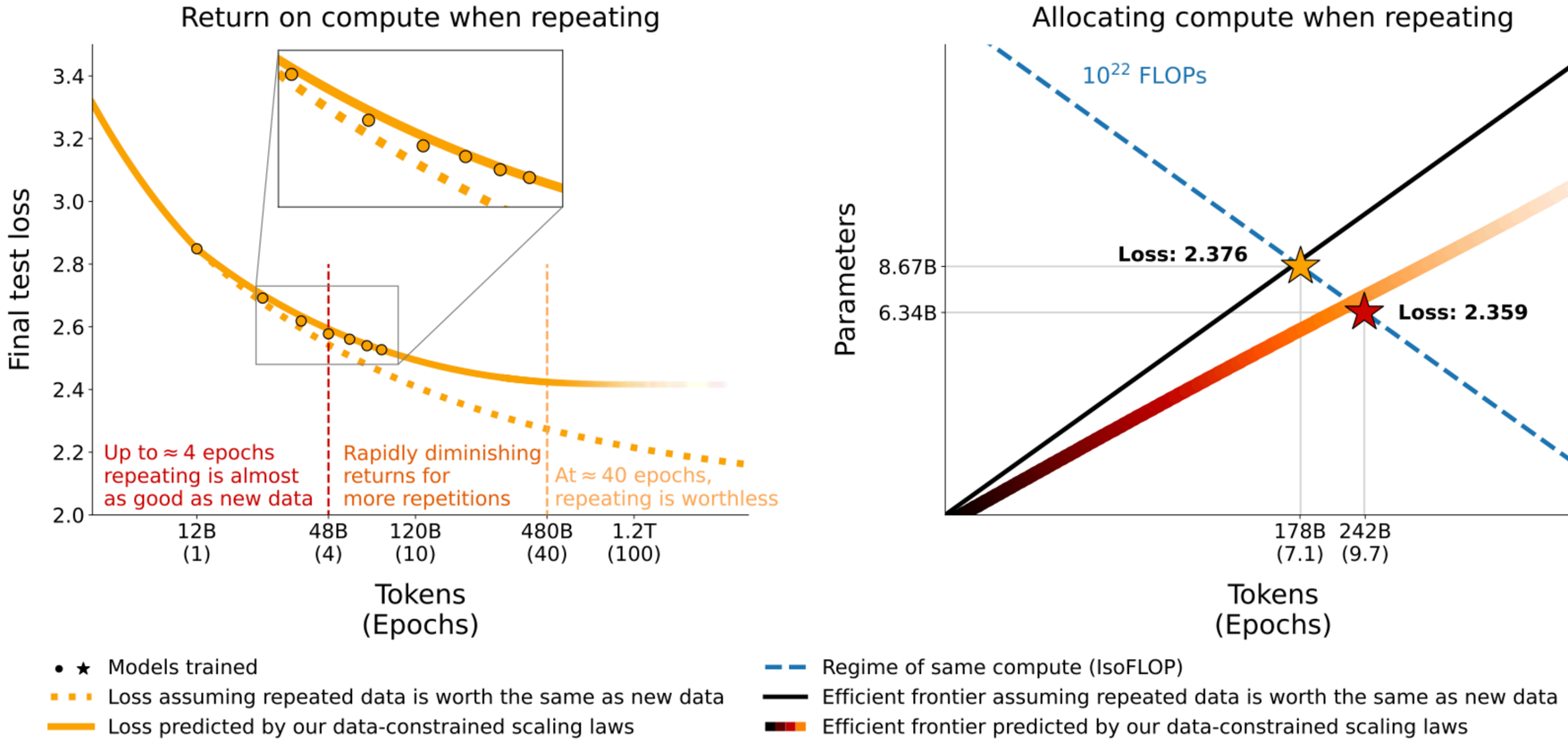
Data Scaling Laws [Kaplan+ 2020]

- So, how does dataset composition affect performance?
  - Dataset composition affects the offset, not the slope



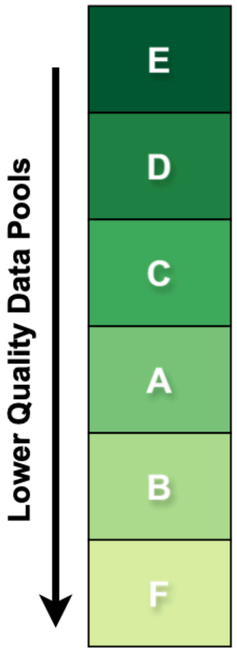
Data Scaling Laws [Muennighoff+ 2023]

- In practice, we have finite data (data constrained scenarios)
  - How does repeating examples affect scaling?

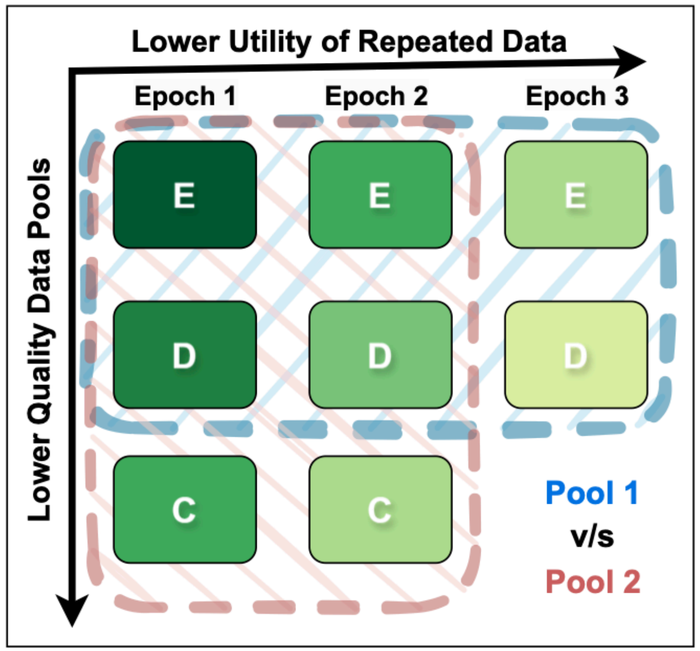


Data Scaling Laws [Goyal+ 2024]

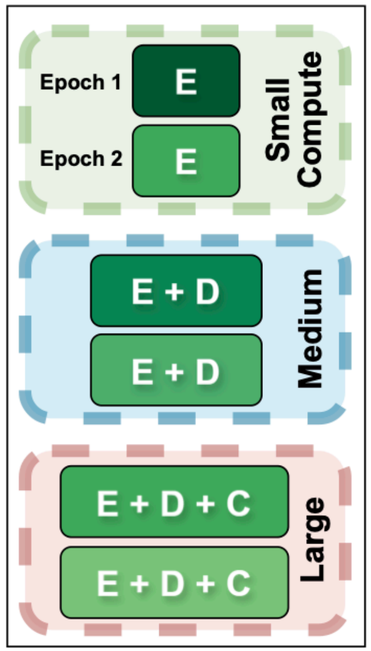
- Data filtering (curation) should be adaptive to scale
  - Less compute: higher filtering / high compute: lower filtering



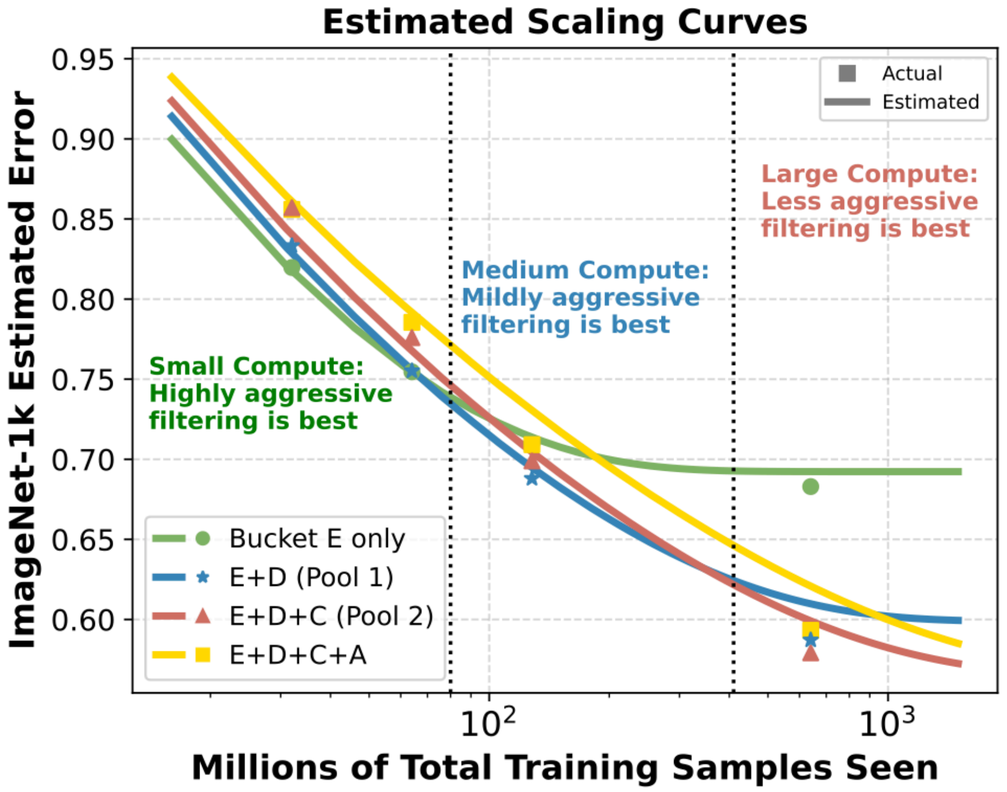
Web Data is Non-Homogenous



Quality-Quantity Tradeoff (QQT) for Data Filtering

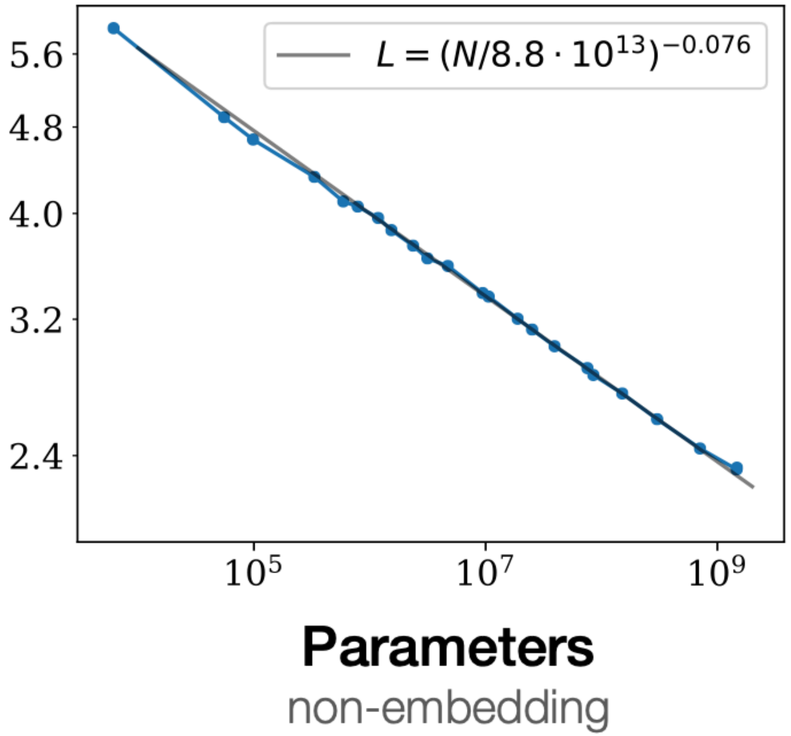


Best Data Pool Changes with Total Compute



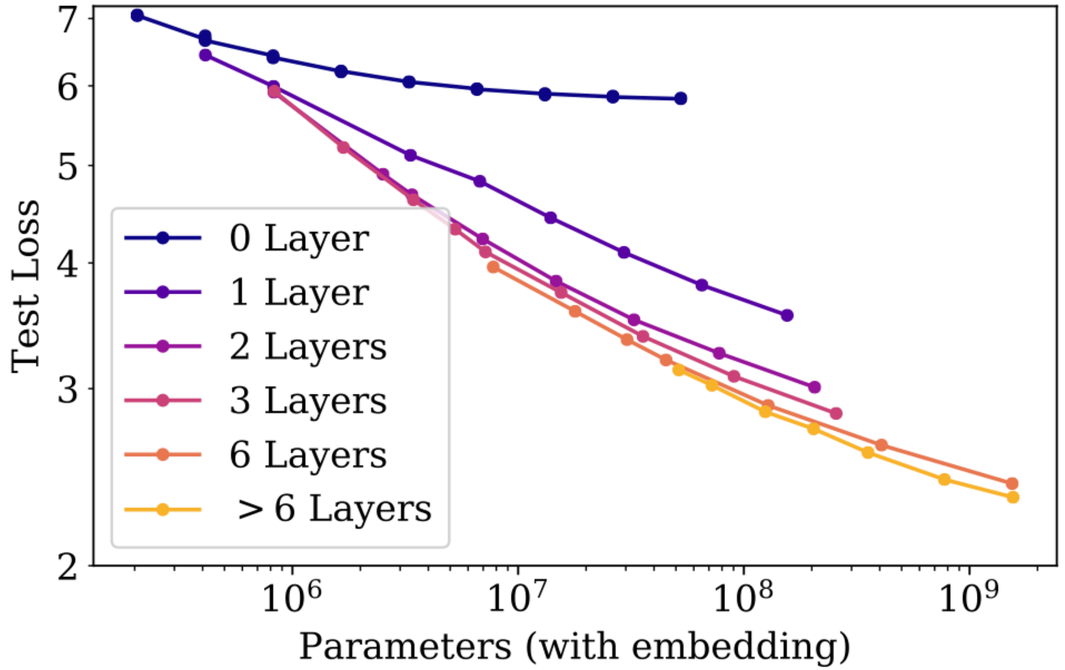
Model Scaling Laws [Kaplan+ 2020]

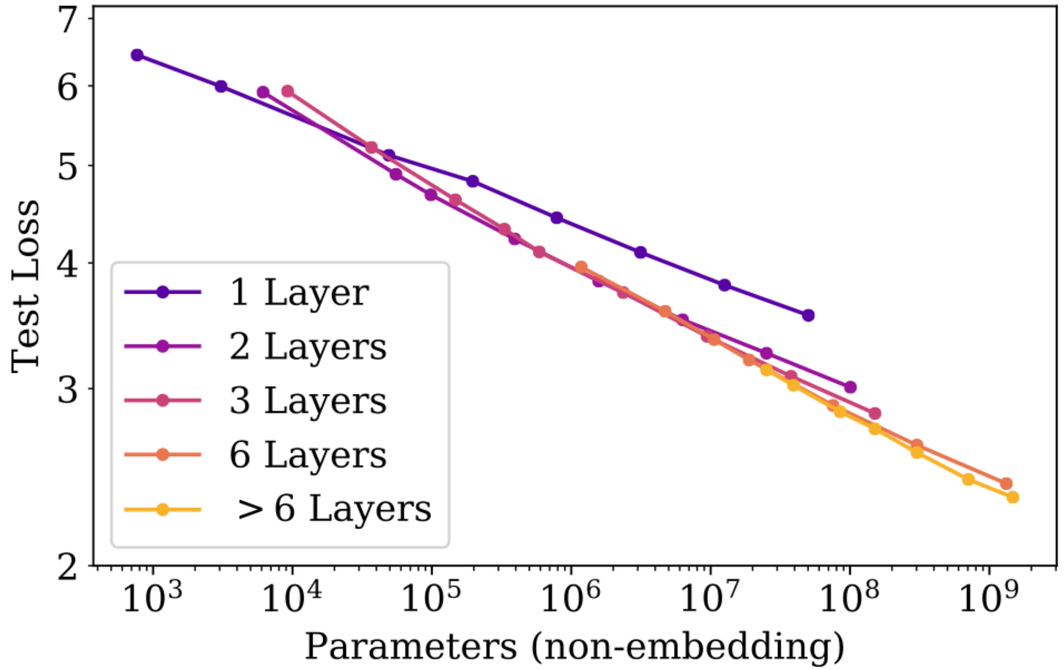
- Performance has (also) a power-law relationship with model size
  - Performance depends strongly on scale, weakly on model shape
  - $L(N) = (N_c/N)^{\alpha_N}$; $\alpha_N \sim 0.076$, $N_c \sim 8.8 \times 10^{13}$ (non-embedding parameters)



Model Scaling Laws [Kaplan+ 2020]

- Note: Be aware that we use non-embedding parameters
  - Embedding parameters is $V \times d_{model}$, which is fixed
  - Model size is small: # embedding params could be dominant
  - Model size is big: # embedding params is negligible





Recap

- Remarkably linear relationship between log-data size and log-error
- Remarkably linear relationship between log-model size and log-error

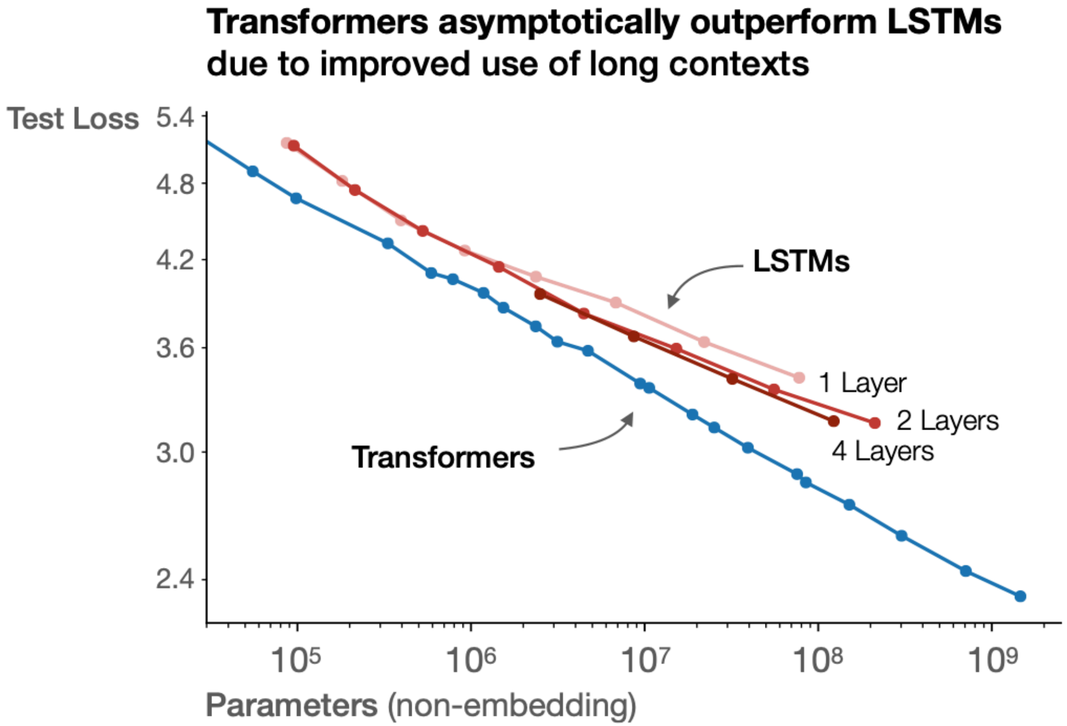
Scaling Laws for Model Engineering

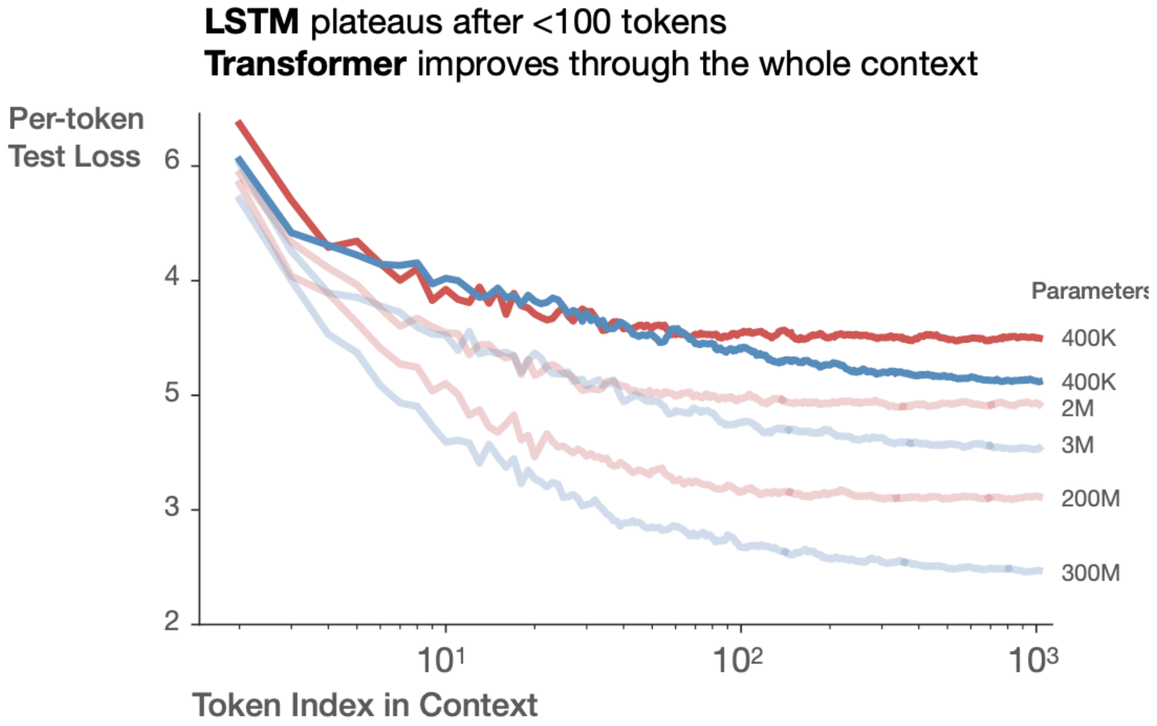
Hyperparameter Questions

- We'll consider some of these choices in the context of [Kaplan+ 2020]
  - Model architecture
  - Optimizer
  - Aspect ratio / depth
  - MoE
  - Batch size

Architecture: Transformer vs. LSTMs [Kaplan+ 2020]

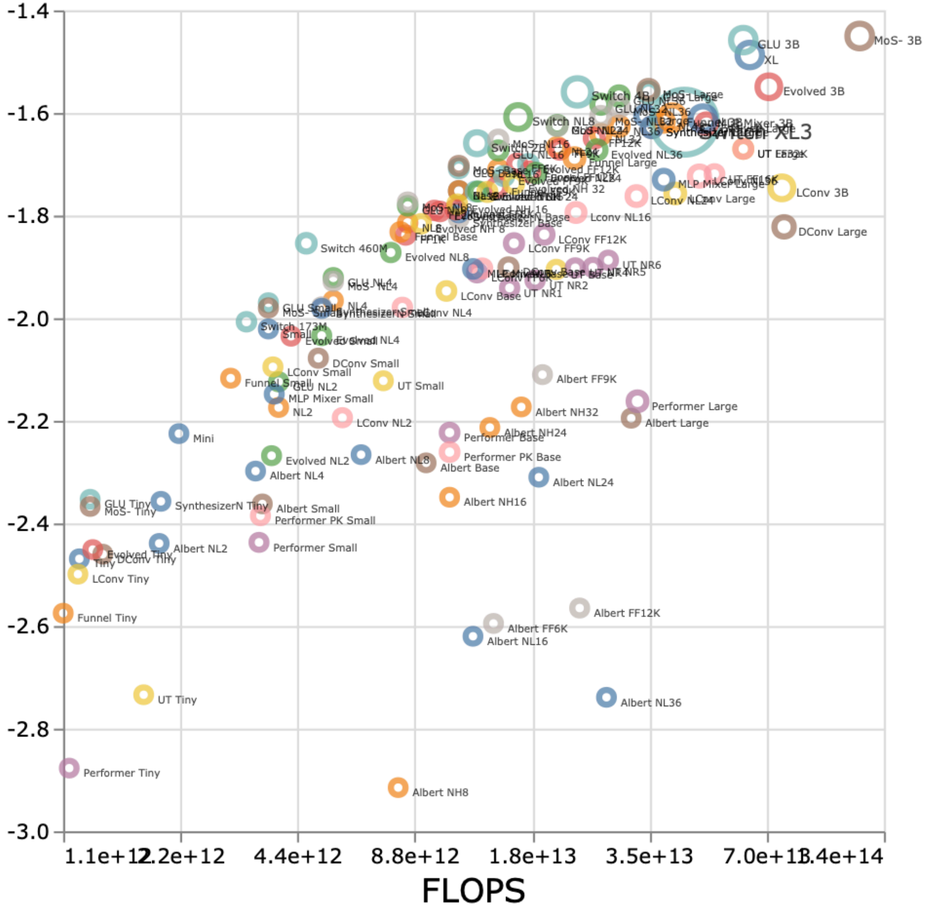
- Q: Are transformers better than LSTMs?
  - A1: Brute force way: spend tens of millions to train an LSTM GPT-3
  - A2: Scaling law way:





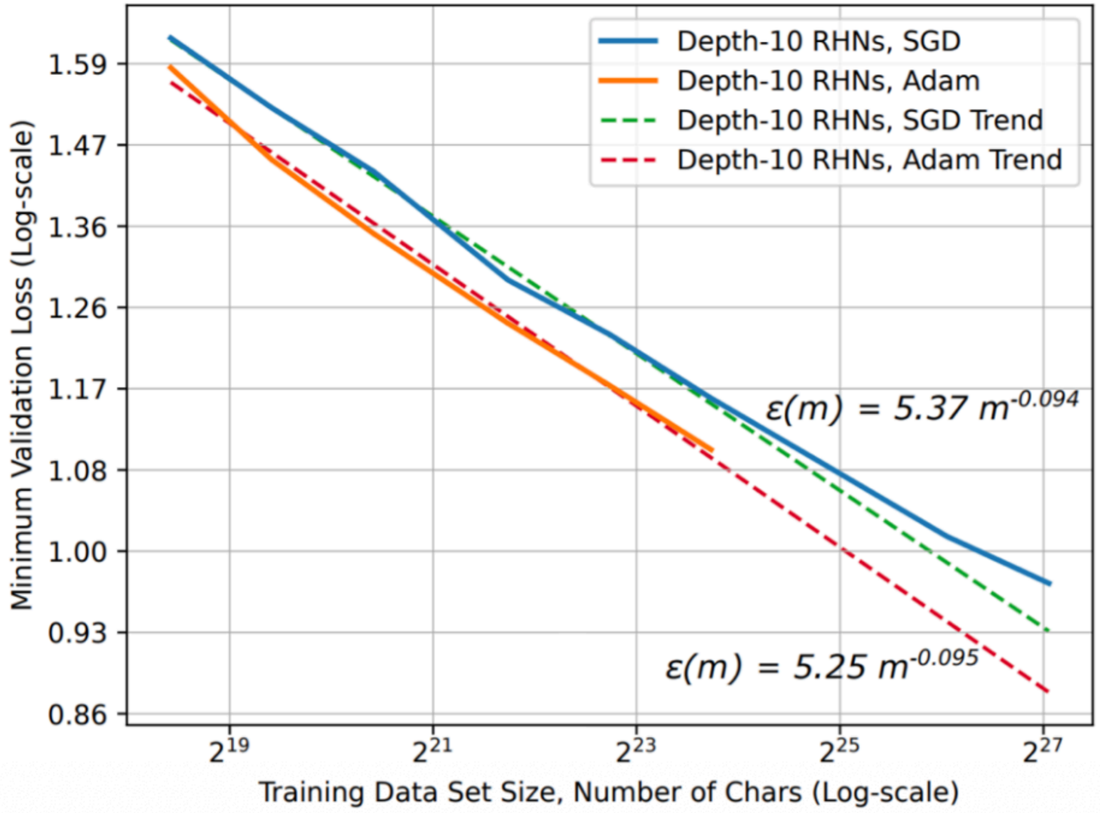
Architecture: Transformer vs. LSTMs

- Cross-architecture scaling in [Tay+ 2022]


-1.4
-1.6
-1.8
-2.0
-2.2
-2.4
-2.6
-2.8
-3.0
1.1e+12
2.2e+12
4.4e+12
8.8e+12
1.8e+13
3.5e+13
7.0e+13
1.4e+14
FLOPS
MoS- 3B
GLU 3B
XL
Evolved 3B
LConv 3B
DConv Large
Switch NL8
Switch 460M
Evolved NL8
GLU NL4
LConv Base
UT NR1
UT NR2
LConv FF12K
LConv FF9K
Albert FF9K
Performer Large
Albert Large
Albert NH32
Albert NH24
Performer Base
Performer PK Base
Albert Base
Albert NL24
Albert NH16
Albert FF12K
Albert FF6K
Albert NL16
Albert NL36
Albert NH8
Mini
Evolved NL2
Albert NL8
Albert NL4
GLU Tiny
SynthesizerN Tiny
Albert Small
Performer PK Small
Albert NL2
Performer Small
Evolved Tiny
DConv Tiny
LConv Tiny
Funnel Tiny
UT Tiny
Performer Tiny
DConv Small
Funnel Small
UT Small
MLP Mixer Small
NL2
LConv NL2
Switch 173M
LConv NL16
LConv FF12K
LConv FF9K
MLP Mixer Large
LConv Large
UT NR6
UT NR5
UT NR4
UT NR3
Switch NL16
Switch NL32
Evolved NH 32
Evolved NH 16
Evolved NH 8
Evolved NH 4
Evolved NL36
UT Base
UT Large
Switch XL3
Switch 2B
Switch 4B
Switch 26B
Switch 52B
Switch 104B
Switch 208B
Switch 395B
Switch 790B
Switch 1571B
Switch 3142B
Switch 6284B
Switch 12582B
Switch 25166B
Switch 50332B
Switch 100663B
Switch 201327B
Switch 402654B
Switch 805308B
Switch 1610616B
Switch 3221232B
Switch 6442464B
Switch 12884928B
Switch 25769856B
Switch 51539712B
Switch 103079424B
Switch 206158848B
Switch 412317696B
Switch 824635392B
Switch 1649270784B
Switch 3298541568B
Switch 6597083136B
Switch 13194166272B
Switch 26388332544B
Switch 52776665088B
Switch 105553330176B
Switch 211106660352B
Switch 422213320704B
Switch 844426641408B
Switch 1688853282816B
Switch 3377706565632B
Switch 6755413131264B
Switch 13510826262528B
Switch 27021652525056B
Switch 54043305050112B
Switch 108086610100224B
Switch 216173220200448B
Switch 432346440400896B
Switch 864692880801792B
Switch 1729385761603584B
Switch 3458771523207168B
Switch 6917543046414336B
Switch 13835086092828672B
Switch 27670172185657344B
Switch 55340344371314688B
Switch 110680688742629376B
Switch 221361377485258752B
Switch 442722754970517504B
Switch 885445509941035008B
Switch 1770891019882070016B
Switch 3541782039764140032B
Switch 7083564079528280064B
Switch 14167128159056560128B
Switch 28334256318113120256B
Switch 56668512636226240512B
Switch 113337025272452481024B
Switch 226674050544904962048B
Switch 453348101089809924096B
Switch 906696202179619848192B
Switch 1813392404359239696384B
Switch 3626784808718479392768B
Switch 7253569617436958785536B
Switch 14507139234873917571072B
Switch 29014278469747835142144B
Switch 58028556939495670284288B
Switch 116057113878991340568576B
Switch 232114227757982681137152B
Switch 464228455515965362274304B
Switch 928456911031930724548608B
Switch 1856913822063861449097216B
Switch 3713827644127722898194432B
Switch 7427655288255445796388864B
Switch 14855310576510891592777728B
Switch 29710621153021783185555456B
Switch 59421242306043566371110912B
Switch 118842484612087132742221824B
Switch 237684969224174265484443648B
Switch 475369938448348530968887296B
Switch 950739876896697061937774592B
Switch 1901479753793394123875549184B
Switch 3802959507586788247751098368B
Switch 7605919015173576495502196736B
Switch 15211838030347152991004393472B
Switch 30423676060694305982008786944B
Switch 60847352121388611964017573888B
Switch 121694704242777223928035147776B
Switch 243389408485554447856070295552B
Switch 486778816971108895712140591104B
Switch 973557633942217791424281182208B
Switch 1947115267884435582848562364416B
Switch 3894230535768871165697124728832B
Switch 7788461071537742331394249457664B
Switch 15576922143075484662788498915328B
Switch 31153844286150969325576997830656B
Switch 62307688572301938651153995661312B
Switch 124615377144603877302307991322624B
Switch 249230754289207754604615982645248B
Switch 498461508578415509209231965290496B
Switch 996923017156831018418463930580992B
Switch 1993846034313662036836927861161984B
Switch 3987692068627324073673855722323968B
Switch 7975384137254648147347711444647936B
Switch 15950768274509296294695422889295872B
Switch 31901536549018592589390845778591744B
Switch 63803073098037185178781691557183488B
Switch 127606146196074370357563383114366976B
Switch 255212292392148740715126766228733952B
Switch 510424584784297481430253532457467904B
Switch 1020849169568594962860507064914935808B
Switch 2041698339137189925721014129829871616B
Switch 4083396678274379851442028259659743232B
Switch 8166793356548759702884056519319486464B
Switch 16333586713097519405768113038638972928B
Switch 32667173426195038811536226077277945856B
Switch 65334346852390077623072452154555891712B
Switch 130668693704780155246144904309111783424B
Switch 261337387409560310492289808618223566848B
Switch 522674774819120620984579617236447133696B
Switch 1045349549638241241969159234472894267392B
Switch 2090699099276482483938318468945788534784B
Switch 4181398198552964967876636937891577069568B
Switch 8362796397105929935753273875783154139136B
Switch 16725592794211859871506547751566308278272B
Switch 33451185588423719743013095503132616556544B
Switch 66902371176847439486026191006265233113088B
Switch 133804742353694878972052382012530466226176B
Switch 267609484707389757944104764025060932452352B
Switch 535218969414779515888209528050121864904704B
Switch 1070437938829559031776419056100243729809408B
Switch 2140875877659118063552838112200487459618816B
Switch 4281751755318236127105676224400974919237632B
Switch 8563503510636472254211352448801949838475264B
Switch 17127007021272944508422704897603899676950528B
Switch 34254014042545889016845409795207799353901056B
Switch 68508028085091778033690819590415598707802112B
Switch 137016056170183556067381639180831197415604224B
Switch 274032112340367112134763278361662394831208448B
Switch 548064224680734224269526556723324789662416896B
Switch 1096128449361468448539053113446649579324833792B
Switch 2192256898722936897078106226893299158649667584B
Switch 4384513797445873794156212453786598317299335168B
Switch 8769027594891747588312424907573196634598670336B
Switch 17538055189783495176624849815146393269197340672B
Switch 35076110379566990353249699630292786538394681344B
Switch 70152220759133980706499399260585573076789362688B
Switch 140304441518267961412998798521171146153578725376B
Switch 280608883036535922825997597042342292307157450752B
Switch 561217766073071845651995194084684584614314901504B
Switch 1122435532146143691303990388169369169228629803008B
Switch 2244871064292287382607980776338738338457259606016B
Switch 4489742128584574765215961552677476676914519212032B
Switch 8979484257169149530431923105354953353829038424064B
Switch 17958968514338299060863846210709906707658076848128B
Switch 35917937028676598121727692421419813415316153696256B
Switch 71835874057353196243455384842839626830632307392512B
Switch 143671748114706392486910769685679253661264614785024B
Switch 287343496229412784973821539371358507322529229570048B
Switch 574686992458825569947643078742717014645058459140096B
Switch 1149373984917651139895286157485434029290116918280192B
Switch 2298747969835302279790572314970868058580233836560384B
Switch 4597495939670604559581144629941736117160467673120768B
Switch 9194991879341209119162289259883472234320935346241536B
Switch 18389983758682418238324578519766944468641870692483072B
Switch 36779967517364836476649157039533888937283741384966144B
Switch 73559935034729672953298314079067777874567482769932288B
Switch 147119870069459345906596628158135555749134965539864576B
Switch 294239740138918691813193256316271111498269931079729152B
Switch 588479480277837383626386512632542222996539862159458304B
Switch 1176958960555674767252773025265084445993079724318916608B
Switch 2353917921111349534505546050530168891986159448637833216B
Switch 4707835842222699069011092101060337783972318897275666432B
Switch 9415671684445398138022184202120675567944637794551332864B
Switch 18831343368890796276044368404241351135889275589102665728B
Switch 37662686737781592552088736808482702271778551178205331456B
Switch 75325373475563185104177473616965404543557102356410662912B
Switch 150650746951126370208354947233930809087114204712821325824B
Switch 301301493902252740416709894467861618174228409425642651648B
Switch 602602987804505480833419788935723236348456818851285303296B
Switch 1205205975609010961666839577871446472696913637702570606592B
Switch 2410411951218021923333679155742892945393827275405141213184B
Switch 4820823902436043846667358311485785890787654550810282426368B
Switch 9641647804872087693334716622971571781575309101620564852736B
Switch 19283295609744175386669433245943143563150618203241129705472B
Switch 38566591219488350773338866491886287126301236406482259410944B
Switch 77133182438976701546677732983772574252602472812964518821888B
Switch 154266364877953403093355465967545148505204945625929037643776B
Switch 308532729755906806186710931935090297010409891251858075287552B
Switch 617065459511813612373421863870180594020819782503716150575104B
Switch 1234130919023627224746843727740361188041639565007432301150208B
Switch 2468261838047254449493687455480722376083279130014864602300416B
Switch 4936523676094508898987374910961444752166558260029729204600832B
Switch 9873047352189017797974749821922889504333116520059458409201664B
Switch 19746094704378035595949499643845779008666233040118916818403328B
Switch 39492189408756071191898999287691558017332466080237833636806656B
Switch 78984378817512142383797998575383116034664932160475667273613312B
Switch 157968757635024284767595997150766232069329864320951334547226624B
Switch 315937515270048569535191994301532464138659728641902669094453248B
Switch 631875030540097139070383988603064928277319457283805338188906496B
Switch 1263750061080194278140767977206129856554638914567610676377812992B
Switch 2527500122160388556281535954412259713109277829135221352755625984B
Switch 5055000244320777112563071908824519426218555658270442705511251968B
Switch 10110000488641554225126143817649038852437111316540885411022503936B
Switch 20220000977283108450252287635298077704874222633081770822045007872B
Switch 40440001954566216900504575270596155409748445266163541644090015744B
Switch 80880003909132433801009150541192310819496890532327083288180031488B
Switch 161760007818264867602018301082384621638993781064654166576360062976B
Switch 323520015636529735204036602164769243277987562129308333152720125952B
Switch 647040031273059470408073204329538486555975124258616666305440251904B
Switch 1294080062546118940816146408659076973111950248517233332610880503808B
Switch 2588160125092237881632292817318153946223900497034466665221761007616B
Switch 5176320250184475763264585634636307892447800994068933330443522015232B
Switch 10352640500368951526529171269272615784895601988137866660887044030464B
Switch 20705281000737903053058342538545231569791203976275733321774088060928B
Switch 41410562001475806106116685077090463139582407952551466643548176121856B
Switch 82821124002951612212233370154180926279164815905102933287096352243712B
Switch 165642248005903224424466740308361852558329631810205866574192704487424B
Switch 331284496011806448848933480616723705116659263620411733148385408974848B
Switch 662568992023612897697866961233447410233318527240823466296770817949696B
Switch 1325137984047225795395733922466894820466637054481646932593541635899392B
Switch 2650275968094451590791467844933789640933274108963293865187083271798784B
Switch 5300551936188903181582935689867579281866548217926587730374166543597568B
Switch 10601103872377806363165871379735158563733096435853175460748333087195136B
Switch 21202207744755612726331742759470317127466192871706350921496666174390272B
Switch 42404415489511225452663485518940634254932385743412701842993332348780544B
Switch 84808830979022450905326971037881268509864771486825403685986664697561088B
Switch 169617661958044901810653942075762537019729542973650807371973329395122176B
Switch 339235323916089803621307884151525074039459085947301614743946658790244352B
Switch 678470647832179607242615768303050148078918171894603229487893317580488704B
Switch 1356941295664359214485231536606100296157836343789206458975786635160977408B
Switch 2713882591328718428970463073212200592315672687578412917951573270321954816B
Switch 5427765182657436857940926146424401184631345375156825835903146540643909632B
Switch 10855530365314873715881852292848802369262690750313651671806293081287819264B
Switch 21711060730629747431763704585697604738525381500627303343612586162575638528B
Switch 43422121461259494863527409171395209477050763001254606687225172325151277056B
Switch 86844242922518989727054818342790418954101526002509213374450344650302554112B
Switch 173688485845037979454109636685580837908203052005018426748900689300605108224B
Switch 347376971690075958908219273371161675816406104010036853497801378601210216448B
Switch 694753943380151917816438546742323351632812208020073706995602757202420432896B
Switch 1389507886760303835632877093484646703265624416040147413991205514404840865792B
Switch 2779015773520607671265754186969293406531248832080294827982411028809681731584B
Switch 5558031547041215342531508373938586813062497664160589655964822057619363463168B
Switch 11116063094082430685063016747877173626124995328321179311929644115238726926336B
Switch 22232126188164861370126033495754347252249990656642358623859288230477853672672B
Switch 44464252376329722740252066991508694504499981313284717247718576460955707345344B
Switch 88928504752659445480504133983017389008999962626569434495437152921911414690688B
Switch 177857009505318890961008267966034778017999925253138868990874305843822829381376B
Switch 355714019010637781922016535932069556035999850506277737981748611687645658762752B
Switch 711428038021275563844033071864139112071999701012555475963497223375291317525504B
Switch 1422856076042551127688066143728278224143999402025110951926994446750582635051008B
Switch 2845712152085102255376132287456556448287998804050221903853988893501165270102016B
Switch 5691424304170204510752264574913112896575997608100443807707977787002330540204032B
Switch 11382848608340409021504529149826225793151995216200887615415955574004661080408064B
Switch 22765697216680818043009058299652451586303990432401775230831911148009322160816128B
Switch 45531394433361636086018116599304903172607980864803550461663822296018644321632256B
Switch 91062788866723272172036233198609806345215961729607100923327644592037288643264512B
Switch 182125577733446544344072466397219612690431923459214201846655289184074577286529024B
Switch 364251155466893088688144932794439225380863846918428403693310578368149154573058048B
Switch 728502310933786177376289865588878450761727693836856807386621156736298309146116096B
Switch 1457004621867572354752579731177756901523455387673713614773242313472596618292232192B
Switch 2914009243735144709505159462355513803046910775347427229546484626945193236584464384B
Switch 5828018487470289419010318924711027606093821550694854459092969253890386473168928768B
Switch 11656036974940578838020637849422055212187643101389708918185938507780772946337857536B
Switch 23312073949881157676041275698844110424375286202779417836371877015561545892675715072B
Switch 46624147899762315352082551397688220848750572405558835672743754031123091785351430144B
Switch 93248295799524630704165102795376441697501144811117671345487508062246183570702860288B
Switch 186496591599049261408330205590752883395002289622235342690975016124492367141405720576B
Switch 372993183198098522816660411181505766790004579244470685381950032248984734282811441152B
Switch 745986366396197045633320822363011533580009158488941370763900064497969468565622882304B
Switch 1491972732792394091266641644726023067160018316977882741527800128995938937131245764608B
Switch 2983945465584788182533283289452046134320036633955765483055600257991877874262491529216B
Switch 5967890931169576365066566578904092268640073267911530966111200515983755748524983058432B
Switch 11935781862339152730133133157808184537280146535823061932222401031967511497049966116864B
Switch 23871563724678305460266266315616369074560293071646123864444802063935022994099932233728B
Switch 47743127449356610920532532631232738149120586143292247728889604127870045988199864467456B
Switch 95486254898713221841065065262465476298241172286584495457779208255740091976399728934912B
Switch 190972509797426443682130130524930952596482344573168990915558510481480183952799457869824B
Switch 381945019594852887364260261049861905192964689146337981831117020962960367905598915739648B
Switch 763890039189705774728520522099723810385929378292675963662234041925920735811197831479296B
Switch 1527780078379411549457041044199447620771858756585351927324468083851841471622395662958592B
Switch 3055560156758823098914082088398895241543717513170703854648936167703682943244791325917184B
Switch 6111120313517646197828164176797790483087435026341407709297872335407365886489582651834368B
Switch 12222240627035292395656328353595580966174870052682815418595744670814731772979165303668736B
Switch 24444481254070584791312656707191161932349740105365630837191489341629463545958330607337472B
Switch 48888962508141169582625313414382323864699480210731261674382978683258927091916661214674944B
Switch 97777925016282339165250626828764647729398960421462523348765957366517854183833322429349888B
Switch 195555850032564678330501253657529295458797920842925046697531914733035708367666844858699776B
Switch 391111700065129356661002507315058590917595841685850093395063829466071416735333689717399552B
Switch 782223400130258713322005014630117181835191683371700186790127658932142833470667379434799104B
Switch 1564446800260517426644010029260234367670383366743400373580255317864285666941334758869598208B
Switch 3128893600521034853288020058520468735340766733486800747160510635728571333882669517739196416B
Switch 6257787201042069706576040117040937470681533466973601494321021271457142667765339035478392832B
Switch 12515574402084139413152080234081874941363066933947202988642042542914285335530678070956785664B
Switch 25031148804168278826304160468163749882726133867894405977284085085828570671061356141913571328B
Switch 50062297608336557652608320936327499765452267735788811954568170171657141342122712283827142656B
Switch 100124595216673115305216641872654999530904535471577623909136340343314282684245424567654285312B
Switch 200249190433346230610433283745309999061809070943155247818272680686628565368490849135308570624B
Switch 400498380866692461220866567490619998123618141886310495636545361373257130736981698270617141248B
Switch 800996761733384922441733134981239996247236283772620991273090722746514261473963396541234282496B
Switch 1601993523466769844883466269962479992494472567545241982546181445493028522947926793082468564992B
Switch 3203987046933539689766932539924959984988945135090483965092362890986057045895853586164937129984B
Switch 6407974093867079379533865079849919969977890270180967930184725781972114091791707172329874259968B
Switch 12815948187734158759067730159699839939955780540361935860369451563944228183583414344659748519936B
Switch 25631896375468317518135460319399679879911561080723871720738903127888456367166828689319497039872B
Switch 51263792750936635036270920638799359759823122161447743441477806255776912734333657378638994079744B
Switch 102527585501873270072541841277598719519646244322895486882955612511553825468667314757277988159488B
Switch 205055171003746540145083682555197439039292488645790973765911225023107650937334629514555976318976B
Switch 410110342007493080290167365110394878078584977291581947531822450046215301874669259029111952637952B
Switch 820220684014986160580334730220789756157169954583163895063644900092430603749338518058223905275904B
Switch 1640441368029972321160669460441579512314339909166327790127289800184861207498677036116447810551808B
Switch 3280882736059944642321338920883159024628679818332655580254579600369722414997354072232895621103616B
Switch 6561765472119889284642677841766318049257359636665311160509159200739444829994708144465791242207232B
Switch 13123530944239778569285355683532636098514719273330622321018318401478889659989416288931582484414464B
Switch 26247061888479557138570711367065272197029438546661244642036636802957779319978832577863164968828928B
Switch 52494123776959114277141422734130544394058877093322489284073273605915558639957665155726329937657856B
Switch 104988247553918228554282845468261088788117754186644978568146547211831117279915330311452659875315712B
Switch 209976495107836457108565690936522177576235508373289957136293094423662234559830660622905319750631424B
Switch 419952990215672914217131381873044355152471016746579914272586188847324469119661321245810639501262848B
Switch 839905980431345828434262763746088710304942033493159828545172377694648938239322642491621279002525696B
Switch 1679811960862691656868525527492177420609884066986319657090344755389297876478645284983242558005051392B
Switch 3359623921725383313737051054984354841219768133972639314180689510778595752957290569966485116010102784B
Switch 6719247843450766627474102109968709682439536267945278628361379021557191505914581139932970232020205568B
Switch 13438495686901533254948204219937419364879072535890557256722758043114383011829162279865940464040411136B
Switch 26876991373803066509896408439874838729758145071781114513445516086228766023658324559731880928080822272B
Switch 53753982747606133019792816879749677459516290143562229026891032172457532047316649119463761856161644544B
Switch 107507965495212266039585633759499354919032580287124458053782064344915064094633298238927523712323289088B
Switch 215015930990424532079171267518998709838065160574248916107564128689830128189266596477855047424646578176B
Switch 430031861980849064158342535037997419676130321148497832215128257379660256378533192955710094849293156352B
Switch 860063723961698128316685070075994839352260642296995664430256514759320512757066385911420189698586312704B
Switch 1720127447923396256633370140151989678704521284593991328860513029518641025514132771822840379397172625408B
Switch 3440254895846792513266740280303979357409042569187982657721026059037282051028265543645680758794345250816B
Switch 6880509791693585026533480560607958714818085138375965315442052118074564102056531087291361517588690501632B
Switch 13761019583387170053066961121215917429636170276751930630884104236149128204113062174582723035177380103264B
Switch 27522039166774340106133922242431834859272340553503861261768208472298256408226124349165446070354760206528B
Switch 55044078333548680212267844484863669718544681107007722523536416944596512816452248698330892140709520413056B
Switch 110088156667097360424535688969727339437089362214015445047072833889193025632904497396661784281419040826112B
Switch 220176313334194720849071377939454678874178724428030890094145667778386051265808994793323568562838081652224B
Switch 440352626668389441698142755878909357748357448856061780188291335556772102531617989586647137125676163304448B
Switch 880705253336778883396285511757818715496714897712123560376582671113544205063235979173294274251352326608896B
Switch 1761410506673557766792571023515637430993429795424247120753165342227088410126471958346588548502704653217792B
Switch 3522821013347115533585142047031274861986859590848494241506330684454176820252943916693177097005409306435584B
Switch 7045642026694231067170284094062549723973719181696988483012661368908353640505887833386354194010818612871168B
Switch 14091284053388462134340568188125099447947438363393976966025322737816707281011775666772708388021637225742336B
Switch 28182568106776924268681136376250198895894876726787953932050645475633414562023551333545416776043274451484672B
Switch 56365136213553848537362272752500397791789753453575907864101290951267228924047102667090833552086548902969344B
Switch 112730272427107697074724545505000795583579506907151815728202581902534457848094205334181667104173097805938688B
Switch 225460544854215394149449091010001591167159013814303631456405163805068915696188410668363334208346195611877376B
Switch 450921089708430788298898182020003182334318027628607262912810327610137831392376821336726668416692391223754752B
Switch 901842179416861576597796364040006364668636055257214525825620655220275662784753642673453336833384782447509504B
Switch 1803684358833723153195592728080012729337272110514429051651241310440551325569507285346906672667768965564895008B
Switch 3607368717667446306391185456160025458674544221028858103302482620881102651139014570693813345335537931129790016B
Switch 7214737435334892612782370912320050917349088442057716206604965241762205302278029141387626690671075862259580032B
Switch 14429474870669785225564741824640101834698176884115432413209930483524410604556058282775253381342151724519160064B
Switch 28858949741339570451129483649280203669396353768230864826419860967048821209112116565550506762684303449038320128B
Switch 57717899482679140902258967298560407338792707536461729652839721934097642418224233131101013525368606898076640256B
Switch 115435798965358281804517934597120814677585415072923459305679443868195284836448466262202027050737213796153280512B
Switch 230871597930716563609035869194241629355170830145846918611358887736390569672896932524404054101474427592306561024B
Switch 461743195861433127218071738388483258710341660291693837222717775472781139345793865048808108202948855184613122048B
Switch 923486391722866254436143476776966517420683320583387674445435550945562278691587730097616216405897710369226096096B
Switch 1846972783445732508872286953553933034841366641166775348890871101891124557383175460195232432811795420738452192192B
Switch 3693945566891465017744573907107866069682733282333550697781742203782249114766350920390464865623590841476904384384B
Switch 7387891133782930035489147814215732139365466564667101395563484407564498229532701840780929731247181682953808768768B
Switch 14775782267565860070978295628431464278730933129334202791126968815128996459065403681561859462494363365907617537536B
Switch 29551564535131720141956591256862928557461866258668405582253937630257992918130807363123718924988726731815235075072B
Switch 59103129070263440283913182513725857114923732517336811164507875261514985836261614726247437849977453463630470150144B
Switch 118206258140526880567826365027451714229847465034673622329015750523029971672523229452494875699954906927260940300288B
Switch 236412516281053761135652730054903428459694930069347244658031501046059943345046458904989751399909813854521880600576B
Switch 472825032562107522271305460109806856919389860138694489316063002092119886690092917809979502799819627709043761201152B
Switch 945650065124215044542610920219613713838779720277388978632126004184239773380185835619959005599639255418087522402304B
Switch 1891300130248430089085221840439227427677559440554777957264252008368479546760371671239918011199278510836175044804608B
Switch 3782600260496860178170443680878454855355118881109555914528504016736959093520743342479836022398557021672350089609216B
Switch 7565200520993720356340887361756909710710237762219111829057008033473918187041486684959672044797114043344700179218432B
Switch 15130401041987440712681774723513819421420475524438223658114016066947836374082973369919344089594228086689400358436864B
Switch 30260802083974881425363549447027638842840951048876447316228032133895672748165946739838688179188456173378800716873728B
Switch 60521604167949762850727098894055277685681902097752894632456064267791345496331893479677376358376912346757601433747456B
Switch 121043208335899525701454197788110555371363804195505789264912128535582690992663786959354752716753824693515202867494912B
Switch 242086416671799051402908395576221110742727608391011578529824257071165381985327573918709505433507649387030405734989824B
Switch 484172833343598102805816791152442221485455216782023157059648514142330763970655147837419010867015298774060811469979648B
Switch 968345666687196205611633582304884442970910433564046314119297028284661527941310295674838021734030597548121622939959296B
Switch 1936691333374392411223267164609768885941820867128092628238594056569323055882620591349676043468061195096243245879918592B
Switch 3873382666748784822446534329219537771883641734256185256477188113138646111765241182699352086936122390192486491759837184B
Switch 7746765333497569644893068658439075543767283468512370512954376226277292223530482365398704173872244780384972983519674368B
Switch 15493530666995139289786137316878151087534566937024741025908752452554584447060964730797408347744489560769945967039348736B
Switch 30987061333990278579572274633756302175069133874049482051817504905109168894121929461594816695488979121539891934078697472B
Switch 61974122667980557159144549267512604350138267748098964103635009810218337788243858923189633390977958243079783868157394944B
Switch 123948245335961114318289098535025208700276535496197928207270019620436675576487717846379266781955916486159567736314789888B
Switch 247896490671922228636578197070050417400553070992395856414540039240873351152975435692758533563911832972319135472629579776B
Switch 495792981343844457273156394140100834801106141984791712829080078481746702305950871385517067127823665944638270945159553552B
Switch 991585962687688914546312788280201669602212283969423659258160156963493404611901742771034134255647331889276541890319107104B
Switch 1983171925375377829092625576560403339204424567938847318516320313926986809223803485542068268511294663778553083780638214208B
Switch 3966343850750755658185251153120806678408849135877694637032640627853973618447606971084136537022589327557106167561276428416B
Switch 7932687701501511316370502306241613356817698271755389274065281255707947236895213942168273074045178655114212335122552856832B
Switch 15865375403003022632741004612483226713635396543510778548130562511415894473790427884336546148090357310228424670245105713664B
Switch 31730750806006045265482009224966453427270793087021557096261125022831788947580855768673092296180714620456849340490211427328B
Switch 63461501612012090530964018449932906854541586174043114192522250045663577895161711537346184592361429240913698680980422854656B
Switch 126923003224024181061928036899865813709083172348086228385044500091327155790323423074692369184722858481827397361960845709312B
Switch 253846006448048362123856073799731627418166344696172456770089000182654311580646846149384738369445716963654794723691419418624B
Switch 507692012896096724247712147599463254836332689392344913540178000365308623161293692298769476738891433927309589447382838837248B
Switch 1015384025792193448495424295198926509672665378784689827080356000730617246322587384597538953477782867854619178894765677674496B
Switch 2030768051584386896990848590397853019345330757569379654160712001461234492645174769195077906955565735709238357789531355348992B
Switch 4061536103168773793981697180795706038690661515138759308321424002922468985290349538390155813911131471418476715579062710697984B
Switch 8123072206337547587963394361591412077381323030277518616642848005844937970580699076780311627822262942836953431158125421395968B
Switch 16246144412675095175926788723182824154762646060555037233285696011689875941161398153560623255644525885673906862316250842791936B
Switch 32492288825350190351853577446365648309525292121110074466571392023379751882322796307121246511289051771347813724632501685583872B
Switch 64984577650700380703707154892731296619050584242220148933142784046759503764645592614242493022578103542695627449265003371167744B
Switch 129969155301400761407414309785462593238101168484440297866285568093519007529291185228484986045156207085391254898530006742335488B
Switch 259938310602801522814828619570925186476202336968880595732571136187038015058582370456969972090312414170782509797060013484670976B
Switch 519876621205603045629657239141850372952404673937761191465142272374076030117164740913939944180624828341565019594120026969341952B
Switch 1039753242411206091259314478283700745904809347875522382930284544748152060234329481827879888361249656683130039188240053938683904B
Switch 2079506484822412182518628956567401491809618695751044765860569089496304120468658963655777672498313366260078376480107877367809808B
Switch 4159012969644824365037257913134802983619237391502089531721138178992608240937317927311555344996626732520156752960215754735619616B
Switch 8318025939289648730074515826269605967238474783004179063442276357985216481874635854623110689993253465040313505920431509471239232B
Switch 16636051878579297460149031652539211934476949566008358126884552715970432963749271709246221379986506930080627011840863018942478464B
Switch 33272103757158594920298063305078423868953899132016716253769105431940865927498543418492442759973013860161254023681726037884956928B
Switch 66544207514317189840596126610156847737907798264033432507538210863881731854997086836984885519946027720322508047363452075769913856B
Switch 133088415028634379681192253220313695475815596528066865015076421727763463709994173673969771039892055440645016094726904151539827712B
Switch 266176830057268759362384506440627390951631193056133730030152843455526927419988347347939542079784110881290032189453808303079655424B
Switch 532353660114537518724769012881254781903262386112267460060305686911053854839976694695879084159568221762580064378907616158311310848B
Switch 1064707320229075037449538025762509563806524772224534920120611373822107709679953389391758168319136443525160128757815232316622621696B
Switch 2129414640458150074899076051525019127613049544449069840241222747644215419359906778783516336638272887050320257515630464633245243392B
Switch 4258829280916300149798152103050038255226099088898139680482445495288430838719813557567032673276545774100640515031260929266490486784B
Switch 8517658561832600299596304206100076510452198177796279360964890990576861677439627115134065346553091548201281030062521858532980972573568B
Switch 17035317123665200599192608412200153020904396355592558721929781981153723354879254230268130693106183096402562060125043717065961147137136B
Switch 34070634247330401198385216824400306041808792711185117443859563962307446709758508460536261386212366192805124120250087434131922294274272B
Switch 68141268494660802396770433648800612083617585422370234887719127924614893419517016921072522772424732385610248240500174868263844588548544B
Switch 136282536989321604793540867297601224167235170844740469775438255849229786839034033842145045544849464771220496481000349736527689177097088B
Switch 272565073978643209587081734595202448334470341689480939550876511698459573678068067684290091089698929542440992962000699473055378354194176B
Switch 545130147957286419174163469190404896668940683378961879101753023396919147356136135368580182179397859084881985924001398946110756708388352B
Switch 1090260295914572838348326938380809793337881366757923758203506046793838294712272270737160364358795718169763971848002797892221513416776704B
Switch 2180520591829145676696653876761619586675762733515847516407012093587676589424544541474320728717591436339527943696005595784443026833553408B
Switch 4361041183658291353393307753523239173351525467031695032814024187175353178849089082948641457435182872679055887392011191568886053667106816B
Switch 8722082367316582706786615507046478346703050934063390065628048374350706357698178165897282914870365745358111774784022383137772107334213632B
Switch 17444164734633165413573231014092956693406101868126780131256096748701412715396356331794565829740731490716223549568044766275544214668427264B
Switch 34888329469266330827146462028185913386812203736253560262512193497402825430792712663589131659481462981432447099136089532551088429336854528B
Switch 69776658938532661654292924056371826773624407472507120525024386994805650861585425327178263318962922965862894198272179065102176858673709056B
Switch 139553317877065323308585848112743653547248814945014241050048773989611301723170850654356526637925845931725788396544358130204353717347418112B
Switch 279106635754130646617171696225487307094497629890028482100097547979222603446341701308713053275851691863451576793088716260408707434694836224B
Switch 558213271508261293234343392450974614188995259780056964200195095958445206892683402617426106551703383726903153586177432520817414869389672448B
Switch 1116426543016522586468686784901949228377990519560113928400390191916890413785366805234852213103406767453806307172354865041634829738779344896B
Switch 2232853086033045172937373569803898456755981039120227856800780383833780827570733610469704426206813534907612614344709730083269659477558689792B
Switch 4465706172066090345874747139607796913511962078240455713601560767667561655141467220939408852413627069815225228689419460166539318955117379584B
Switch 8931412344132180691749494279215593827023924156480911427203121535335123310282934441878817704827254139630450457378838920332638637910234759168B
Switch 17862824688264361383498988558431187654047848312961822854406243070670246620565868883757635409654508279260900914757677840665277275820469518336B
Switch 35725649376528722766997977116862375308095696625923645708812486141340493241131737767515270819309016558521801829515355681330554551640939036672B
Switch 71451298753057445533995954233724750616191393251847291417624972282680986482263475535030541638618033117043603659030711362661109103281878073344B
Switch 142902597506114891067991908467449501232382786503694582835249944565361972964526951070061277236066234086007318060061422723322206562756146688B
Switch 285805195012229782135983816934899002464765573007389165670499889130723945929053902140122554472132468172014636120122845446644413125512293376B
Switch 571610390024459564271967633869798004929531146014778331340999778261447891858107804280245108944264936344029272240245690893288826251024586752B
Switch 1143220780048919128543935267739596009859062292029556662681999556522895783716215608560490217888529872688058544480491381786577652502049173504B
Switch 2286441560097838257087870535479192019718124584059113325363999113045791567432431217120980435777059745376117088960982763573155305004098347008B
Switch 4572883120195676514175741070958384039436249168118226650727998226091583134864862434241960871554119490752234177921965527146310610008196694016B
Switch 9145766240391353028351482141916768078872498336236453301455996452183166269729724868483921743108238981504468355843931054292621220016393388032B
Switch 18291532480782706056702964283833536157744996672472906602911992904366332539459449736967843486216477963008936711687862108585242440032786776064B
Switch 36583064961565412113405928567667072315489993344945813205823985808732665078918899473935686972432955926017873423375724217170484880065573552128B
Switch 73166129923130824226811857135334144630979986689891626411647971617465330157837798947871373944865911852035746846751448434341761960131147104256B
Switch 146332259846261648453623714270668289261959973379783252823295943234930660315675597895742747889731823704071493693502896868683523920262294208512B
Switch 292664519692523296907247428541336578523919946759566505646591886469861320631351195791485495779463647408142987387005793737367047840524588417024B
Switch 585329039385046593814494857082673157047839893519133011293183772939722641262702391582970991558927294816285974774011587474734095681049176834048B
Switch 1170658078770093187628989714165346314095679787038266022586367545879445282525404783165941983117854589632571949548023175149468191362098353668096B
Switch 2341316157540186375257979428330692628191359574076532045172735091758890565050809566331883966235709179265143899096046350298936382724196707336192B
Switch 4682632315080372750515958856661385256382719148153064090345470183517781130101619132663767932471418358530287798192092700597872765448393414672384B
Switch 9365264630160745501031917713322770512765438296306128180690940367035562260203238265327535864942836717060575596384185401195745530896786829344768B
Switch 18730529260321491002063835426645541025530876592612256361381880734071124520406476530655071729885673434121151192768370802391491061793573658689536B
Switch 37461058520642982004127670853291082051061753185224512722763761468142249040812953061310143459771346868242302385536741604782982123587147317379072B
Switch 74922117041285964008255341706582164102123506370449025445527522936284498081625906122620286919542693736484604771483209565964247174294634758144B
Switch 149844234082571928016510683413164328204247012740898050891055045872568996163251812245240573839085387472969209542966419131928494348589269516288B
Switch 299688468165143856033021366826328656408494025481796101782110091745137992326503624490481147678170774945938419085932838263856988697178539032576B
Switch 599376936330287712066042733652657312816988050963592203564220183490275984653007248980962295356341549891876838171865676527713977394357078065152B
Switch 1198753872660575424132085467305314625633976101927184407128440366980551969306014497961924590712683099783753676343731353055427954788714156130304B
Switch 2397507745321150848264170934610629251267952203854368814256880733961103938612028995923849181425366199567507352687462706110855909577428312260608B
Switch 4795015490642301696528341869221258502535904407708737628513761467922207877224057991847698362850732399135014705374925412221711819154856624521216B
Switch 9590030981284603393056683738442517005071808815417475257027522935844415754448115983695396725701464798270029410749850824443423638309713249042432B
Switch 19180061962569206786113367476885034010143617630834950514055045871688831508896231967390793451402929596540058821499701648886847276619426498084864B
Switch 38360123925138413572226734953770068020287235261669901028110091743377663017792463934781586902805859193080117642999403297773694553238852996169728B
Switch 76720247850276827144453469907540136040574470523339802056220183486755326035584927869563173805611718386160235285998806595547389106477705992339456B
Switch 153440495700553654288906939815080272081148941046679604112440366973510652071169855739126347611223436772320470571997613191094778212955411984678912B
Switch 306880991401107308577813879630160544162297882093359208224880733947021304142339711478252695222872873544640941143995226382189556425910823969357824B
Switch 613761982802214617155627759260321088324595764186718416449761467894042608284679422956505390445745747089281886287990452764379112851821939714715648B
Switch 1227523965604429234311255518520642176649191528373436832899522935788085216569358845913010780891491494178563772575980905528758225703643879429431296B
Switch 2455047931208858468622511037041284353298383056746873665799045871576170433138717691826021561782982988357127545151961811057516451407287758858862592B
Switch 4910095862417716937245022074082568706596766113493747331598091743152340866277435383652043123565965976714255090303923622115032902814575517717185

Optimizer Choice [Hestness+ 2017]

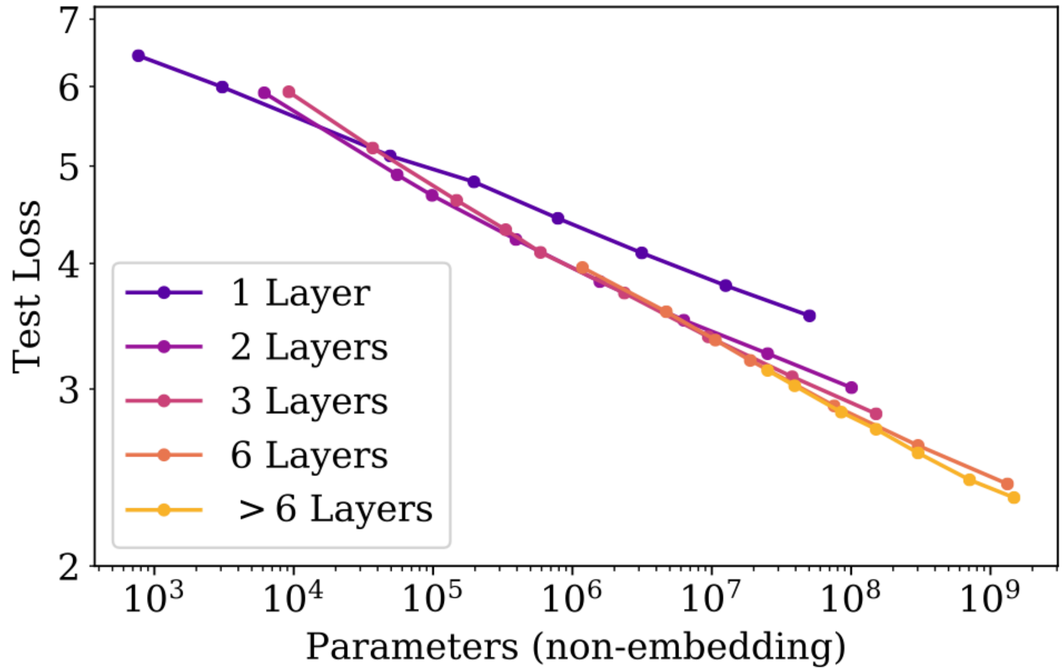
- Q: What about Adam vs. SGD?
  - A: Similar slope, but offsets are different



Note: below results are not Transformer

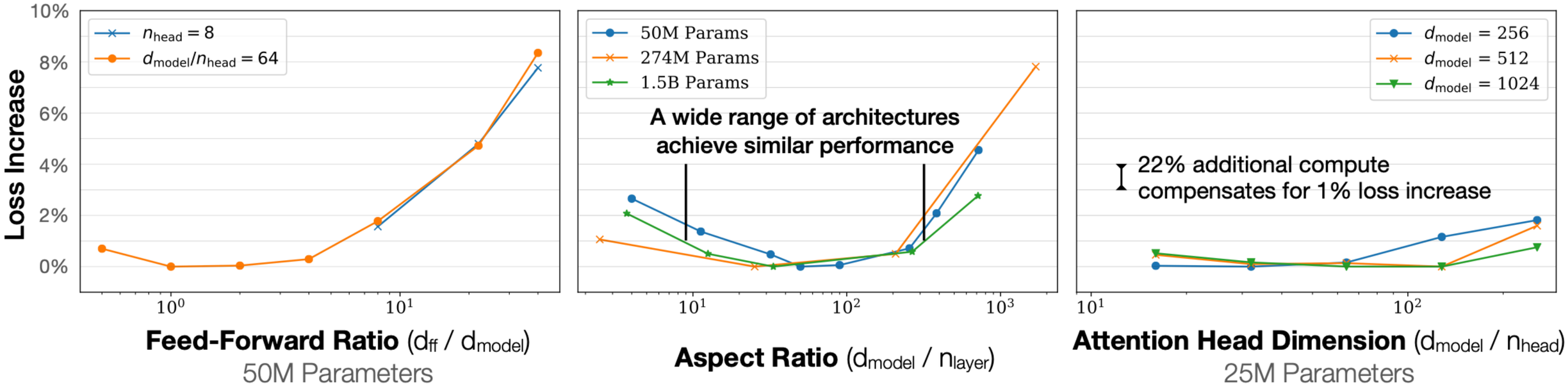
Depth/Width: # of Layers [Kaplan+ 2020]

- Q: Does depth or width make a huge difference?
  - A: 1 vs 2 layers makes a huge difference
  - A: More layers have diminishing returns below 10^7 params



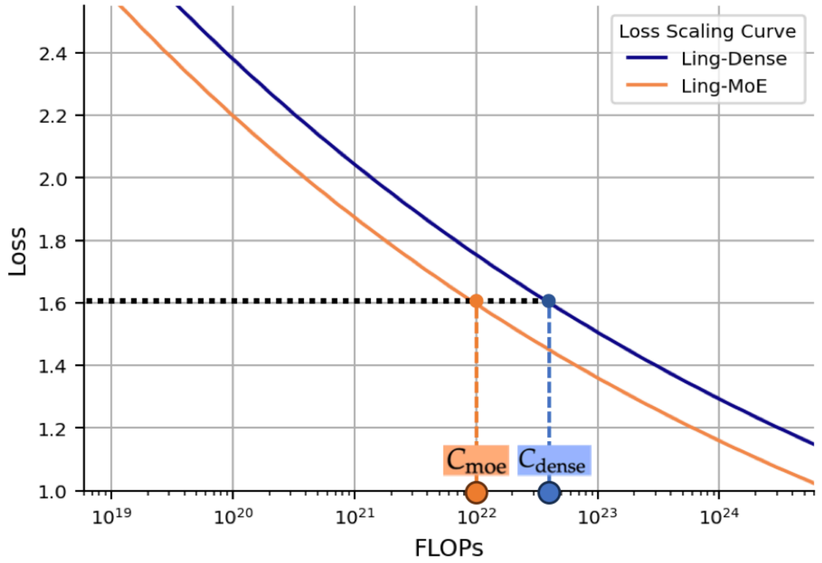
Other Hyperparameters [Kaplan+ 2020]

- Q: Do hyperparameters like the aspect ratio depend on scale?
  - A: Performance depends very mildly on model shape



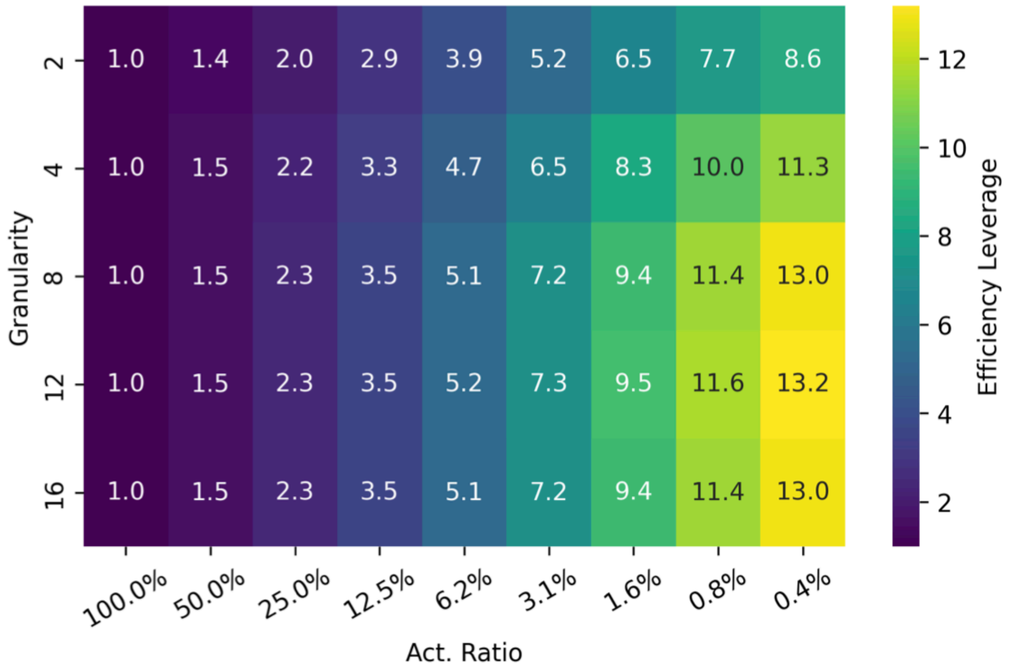
Scaling Laws for MoE [Tian+ 2025]

- MoEs decouples total params from compute cost → hard to predict loss
- Efficiency Leverage: compute efficiency of MoE vs. dense counterpart



$$EL(\mathcal{X}_{\text{MoE}} \mid \mathcal{X}_{\text{Dense}}) = \frac{C_{\text{dense}}}{C_{\text{moe}}}$$

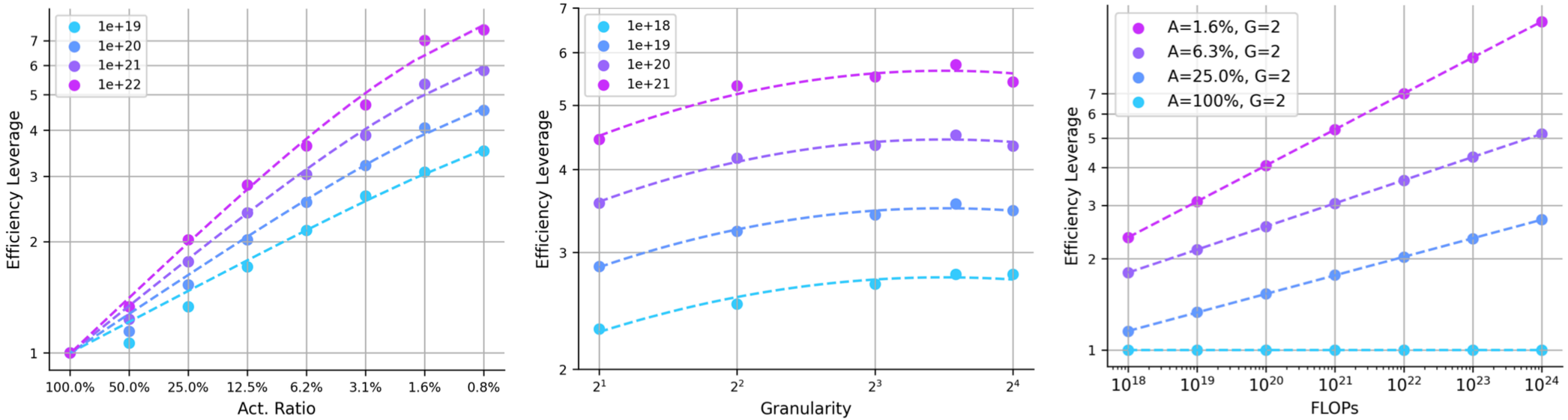
(a) Definition of Efficiency Leverage (EL)



(b) Estimated EL at 1e22 FLOPs

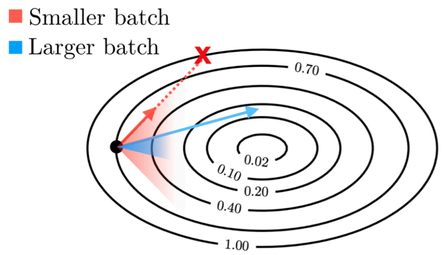
Scaling Laws for MoE [Tian+ 2025]

- Key factors of MoE scaling laws
  - Activation ratio: more sparse, more efficient (monotonically)
  - Expert granularity: there is an optimal granularity (around 12)
  - Compute budget: the bigger model, the effect of MoEs is greater



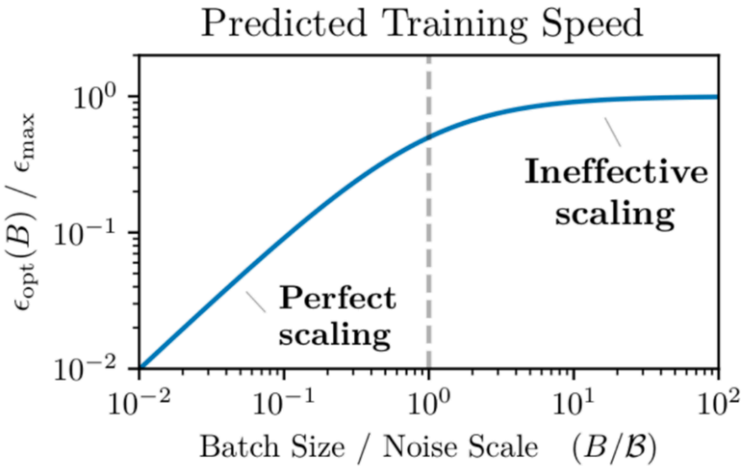
(a) Scaling with Activation Ratio (A) **(b)** Scaling with Granularity (G) **(c)** Scaling with Compute Budget (C)

Batch Size: "Critical Batch Size"



- Batch size has strong diminishing returns past a certain point
  - Small batch size: high-variance gradient (noisy)
  - Large batch size: low-variance gradient
  - Very large batch size: redundant signal; not much gain from it

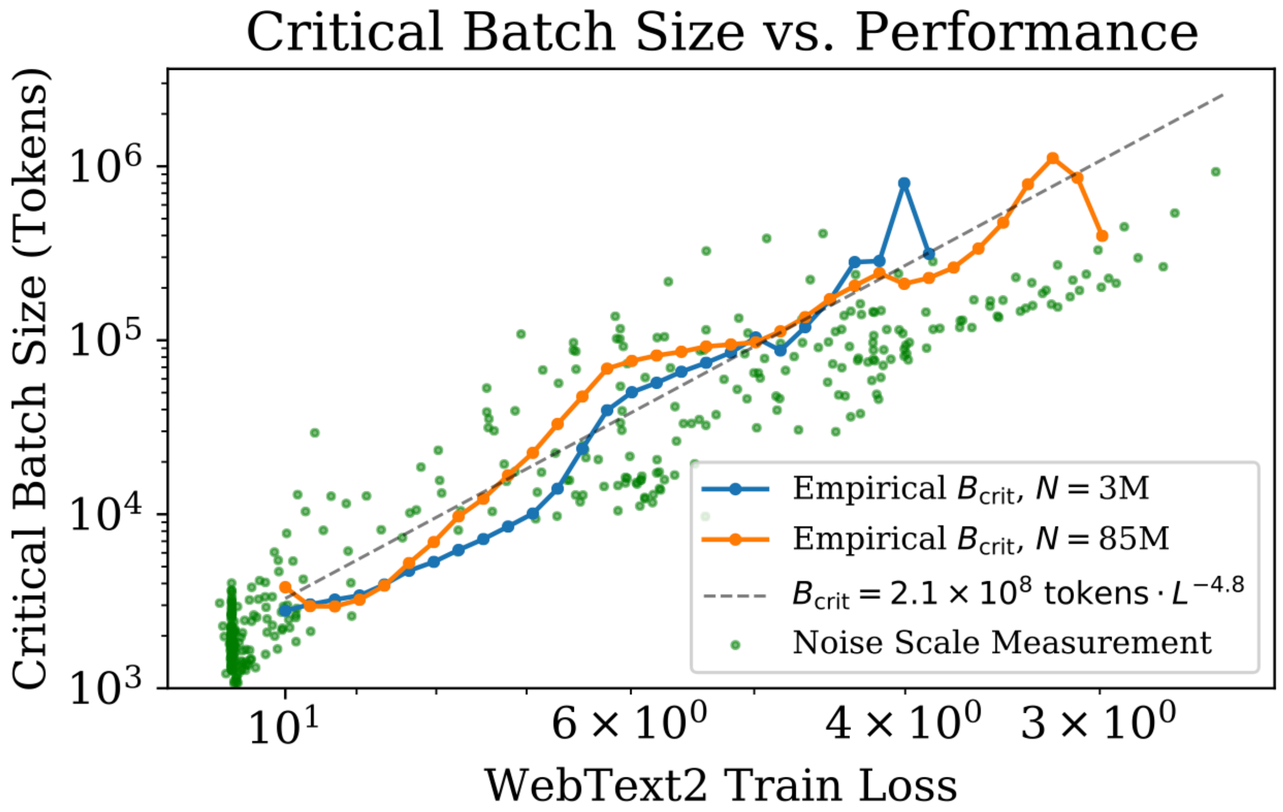
- Critical batch size: $\dfrac{\min(\#\text{ of data for target loss})}{\min(\#\text{ of steps for target loss})}$



  - The Pareto-optimal (i.e. best batch size) that maximizes training speed while maintaining compute efficiency by avoiding the diminishing returns of gradient redundancy

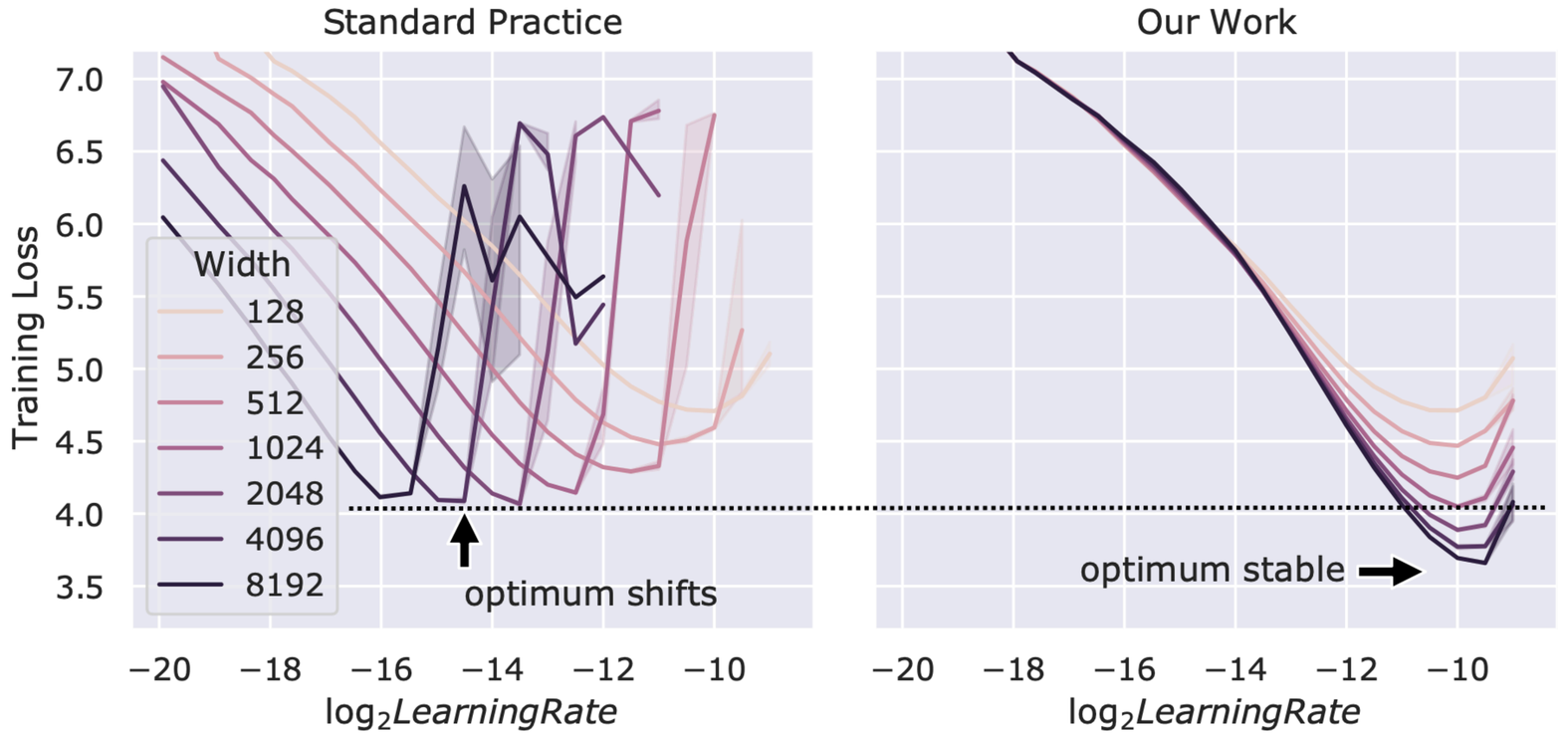
Batch Size: "Critical Batch Size" [Kaplan+ 2020]

- Q: Critical batch size is always the same?
  - A: No! the smaller the "target" loss, the bigger the critical batch size



Learning Rate: μP [Yang+ 2022]

- Q: Should we find optimal learning rate for each scaled-up model?
  - A1: In standard practice, yes 😢
  - A2: μP is scaling-aware initialization and learning rate scaling



Some Takeaways

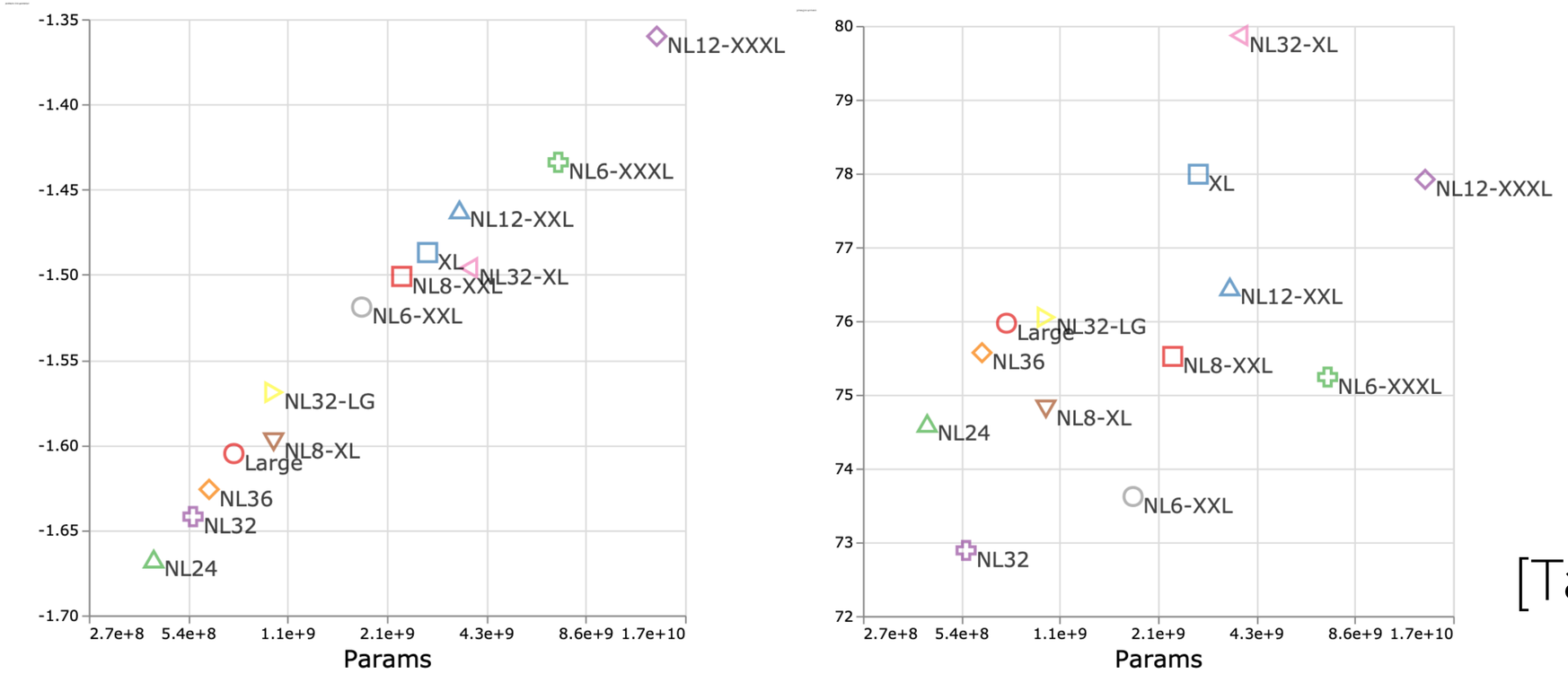
- Effect of HPs on big LLMs can be predicted before training
  - Optimizer choice
  - Model design (depth, aspect ratio, architecture choice, …)

- The scaling law based design procedure
  - Train a few smaller models with a small compute budget
  - Establish my own scaling law
  - Select optimal hyperparams based on the scaling law prediction

- If we don't have scaling laws…
  - We should tune very big models, and waste lots of computing

But Caution…

- Scaling behaviors can differ downstream
  - Thus far, we saw the scaling laws on the losses (i.e. upstream tasks)
  - But downstream scaling can often be much less predictable



[Tay+ 2021]

Compute Allocation

Compute Allocation

- Let's say we can use 1,000 H200 GPUs for a month

- How should we allocate our limited resources?
  - Train models longer vs train bigger models?
  - Collect more data vs get more GPUs?

- Scaling laws can also provide a simple procedure to answer this

Implications of Kaplan's Scaling Law

- Q: Do we need more data or bigger models?
  - A: Clearly, it seems that lots of data is wasted on small models

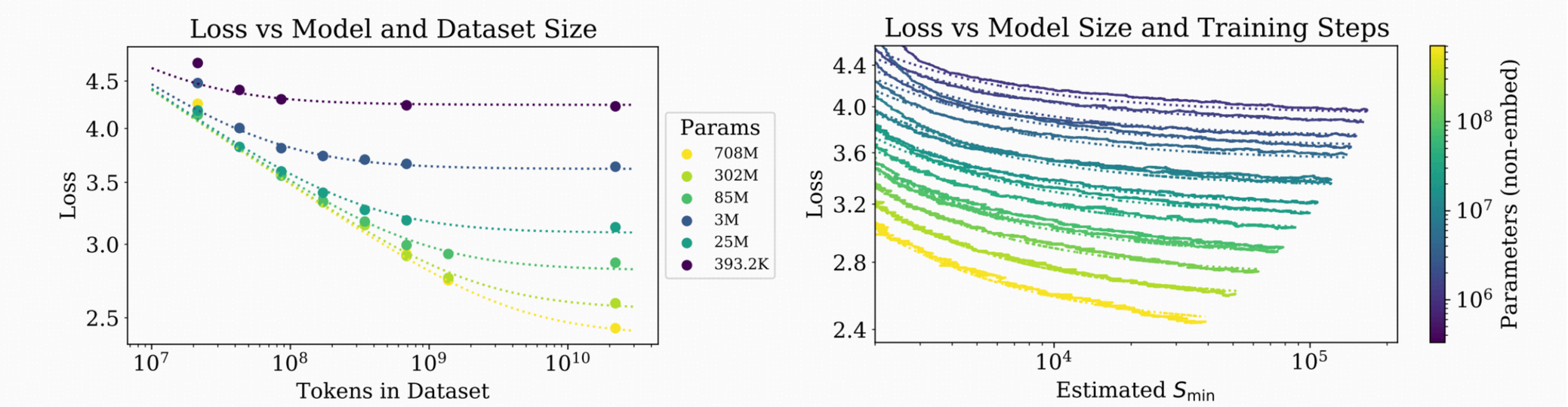
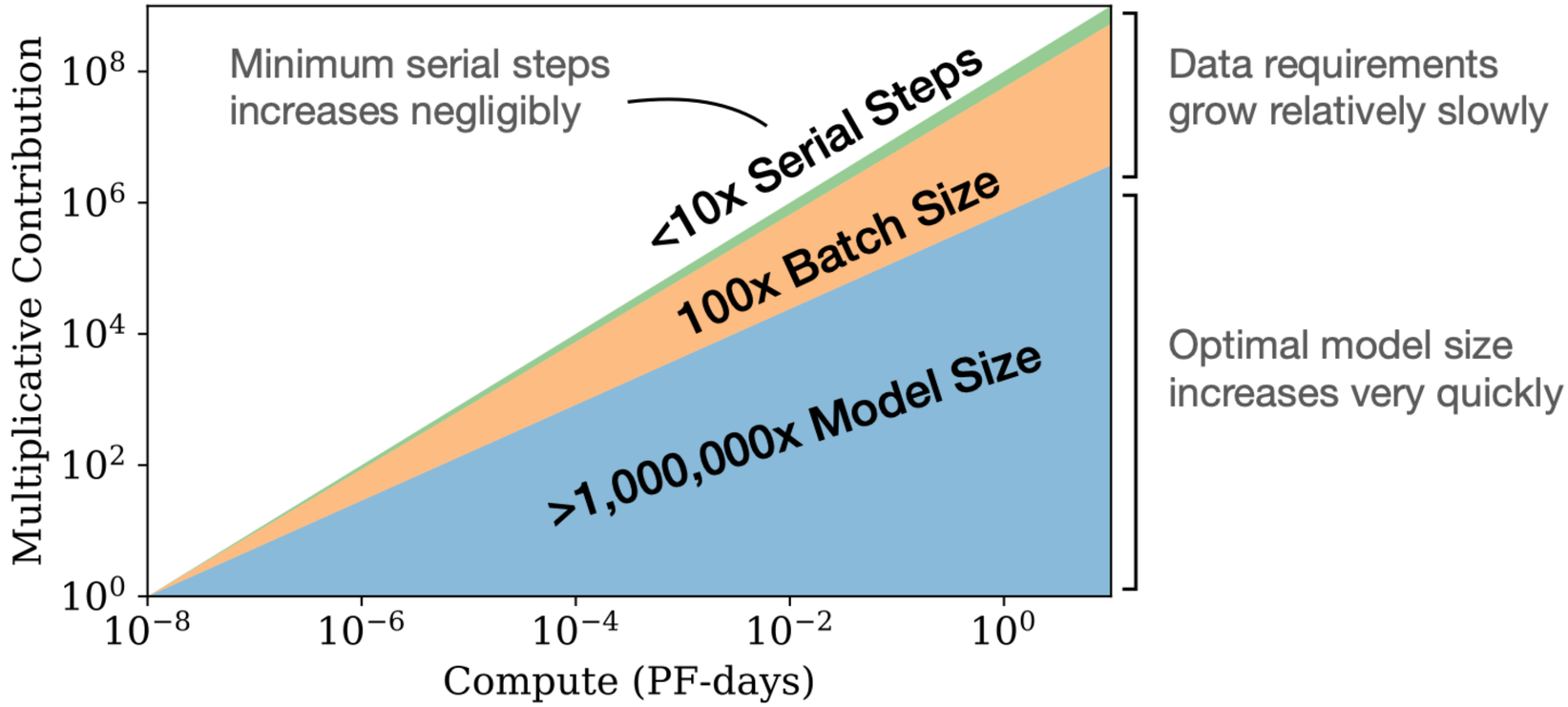


Figure 4 **Left**: The early-stopped test loss $L(N, D)$ varies predictably with the dataset size D and model size N according to Equation (1.5). **Right**: After an initial transient period, learning curves for all model sizes N can be fit with Equation (1.6), which is parameterized in terms of $S_{\min}$, the number of steps when training at large batch size (details in Section 5.1).

[Kaplan+ 2020]

Implications of Kaplan's Scaling Law

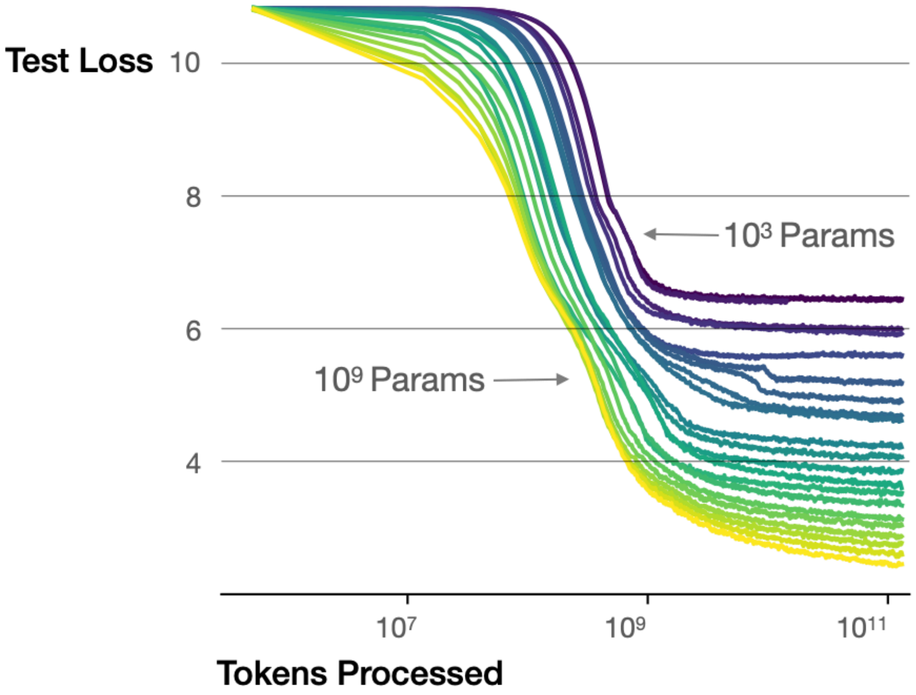
- Q: Do we need more data or bigger models?
  - A: Clearly, it seems that lots of data is wasted on small models



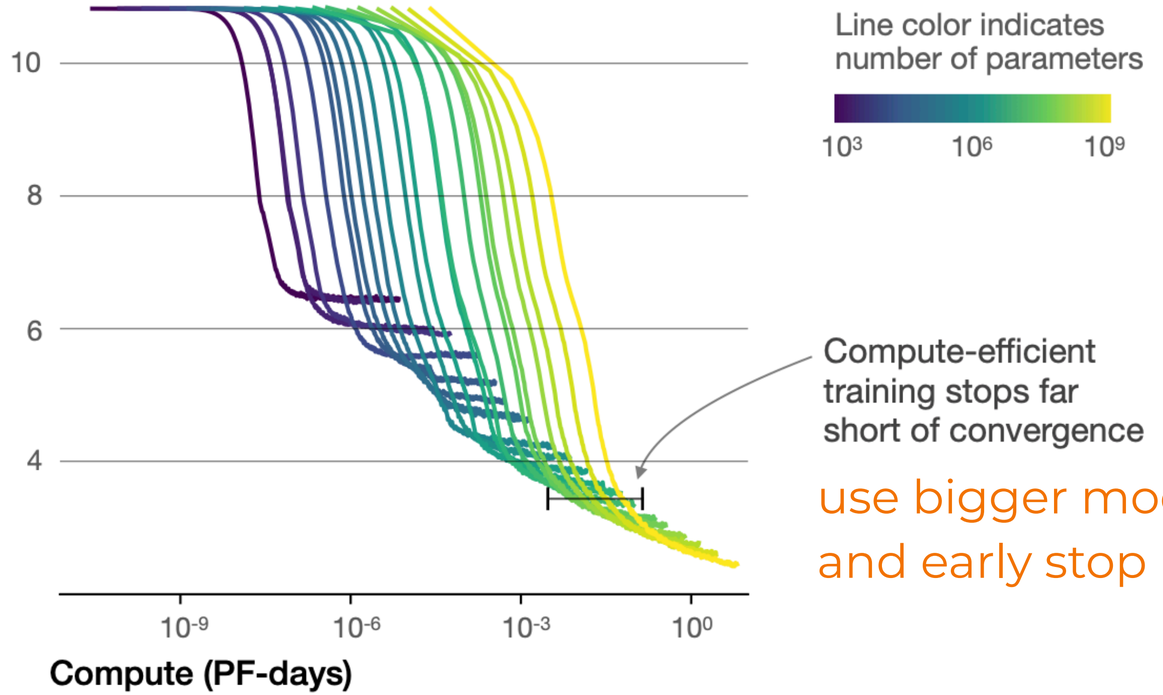
[Kaplan+ 2020]

Training a Bigger Model [Kaplan+ 2020]

Larger models require **fewer samples** to reach the same performance



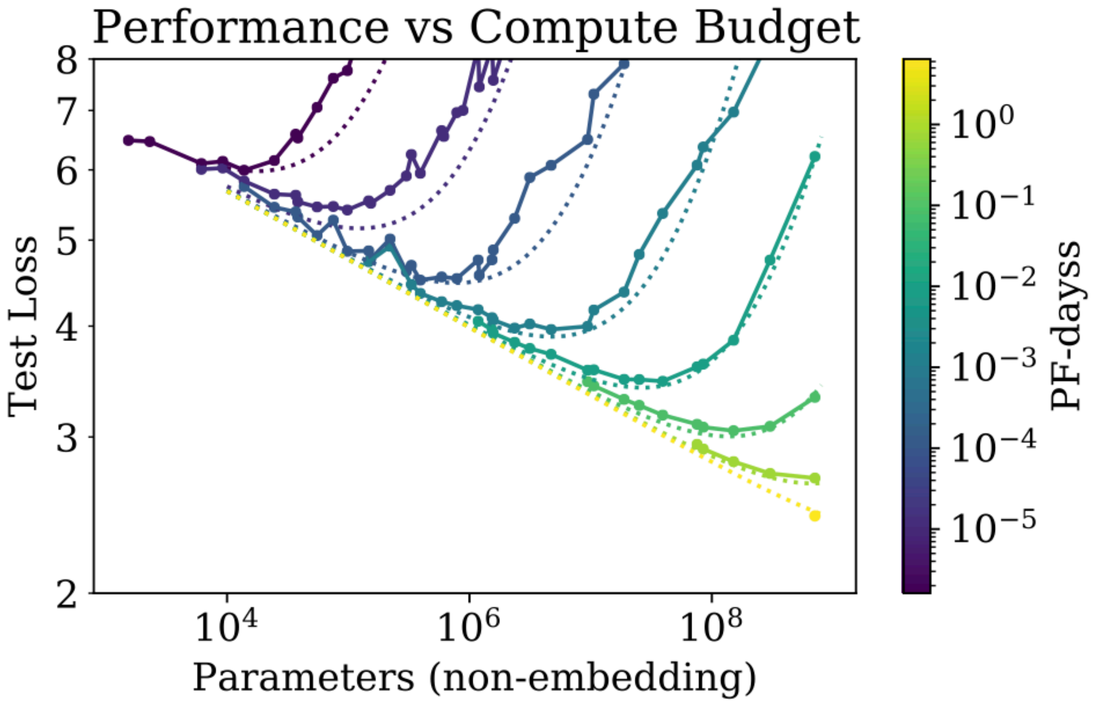
The optimal model size grows smoothly with the loss target and compute budget

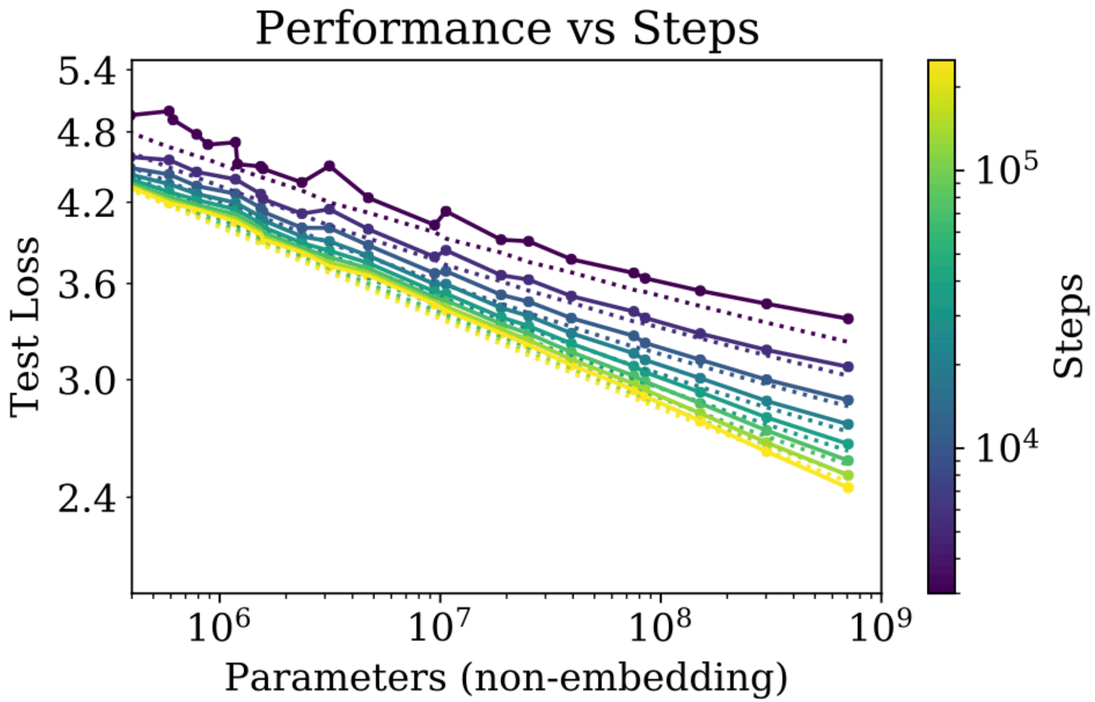


use bigger model, and early stop

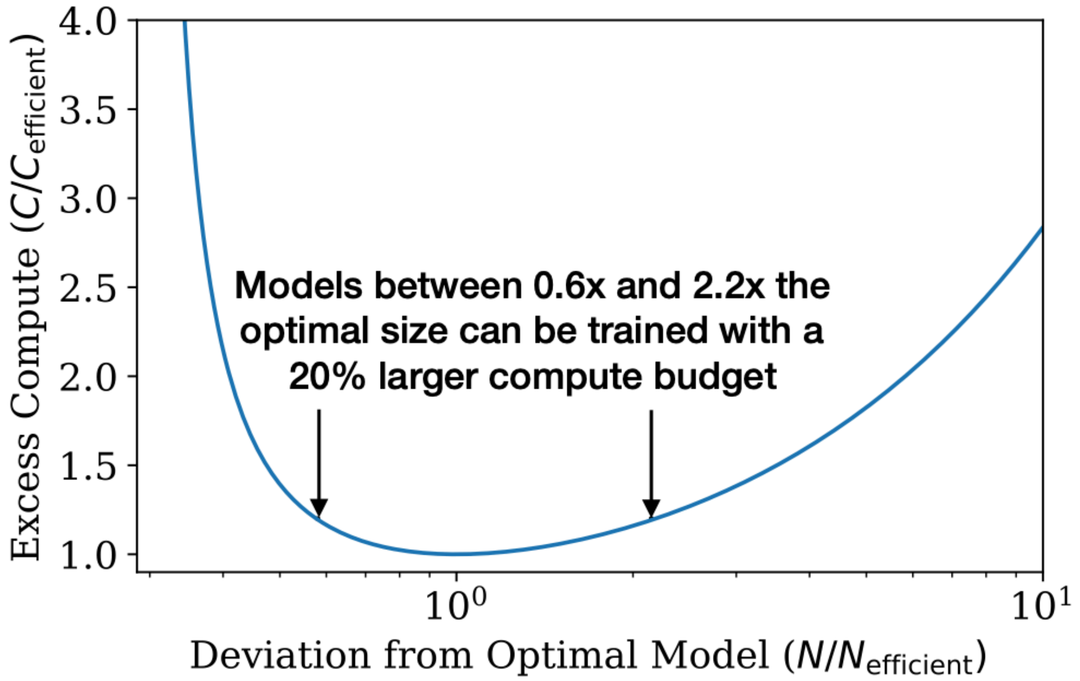
Compute vs. Performance [Kaplan+ 2020]

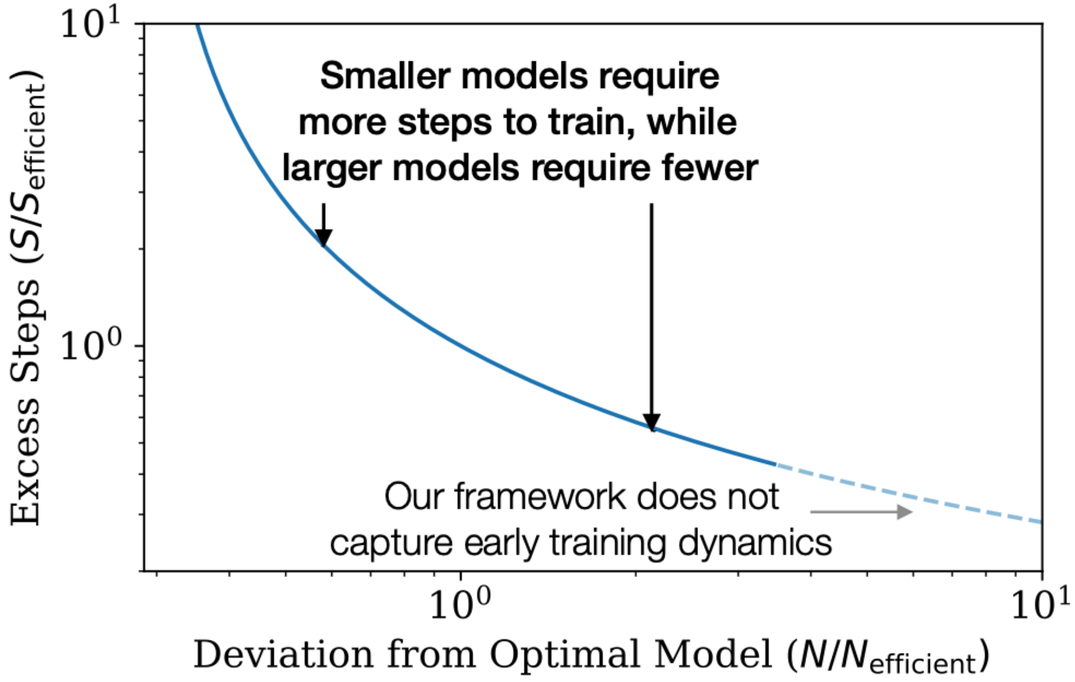
- For a fixed compute budget, there is an optimal model size
  - If we introduce more compute, optimal model size grows





Non-Optimal Model Size? [Kaplan+ 2020]

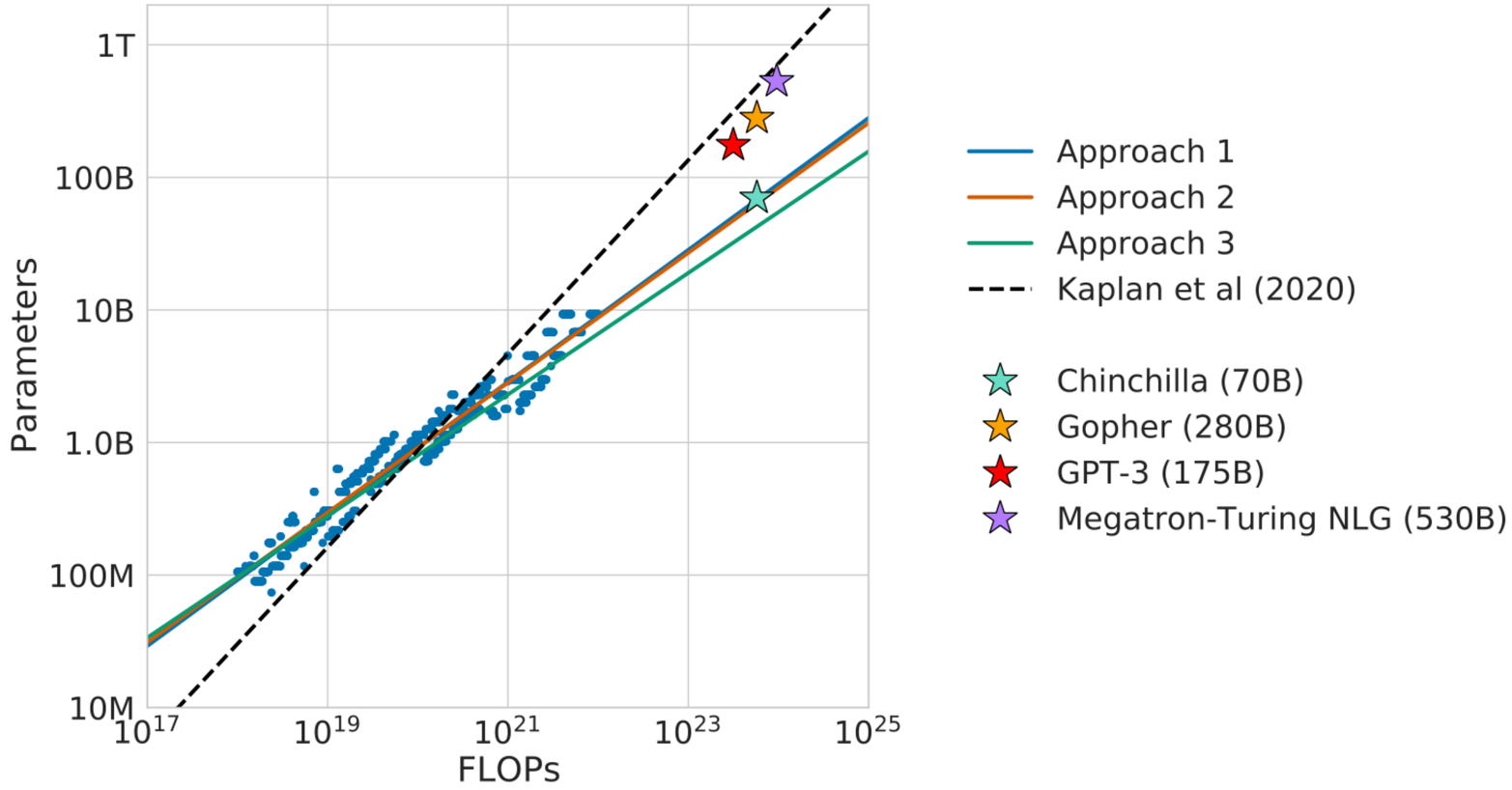




If we have unlimited resources,
train bigger model than the optimal one
with fewer steps

But, "Optimal" Scaling Laws Are Hard to Get

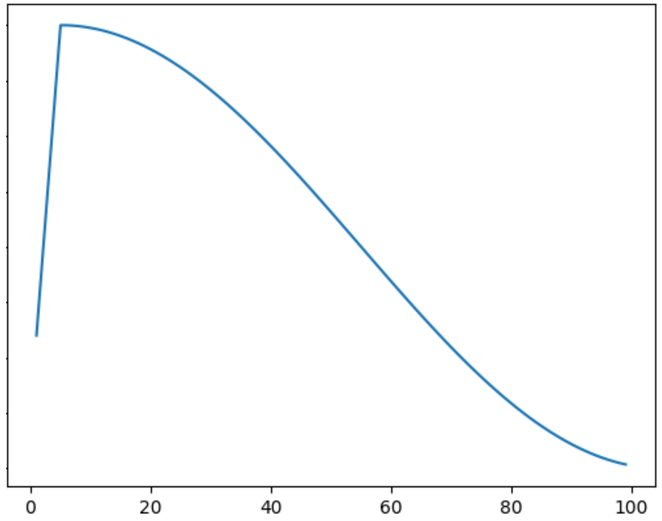
- [Hoffman+ 2022] argue that Kaplan's scaling laws are quite off-fit
  - Kaplan's scaling laws over-estimate the importance of model size



What Makes the Difference?

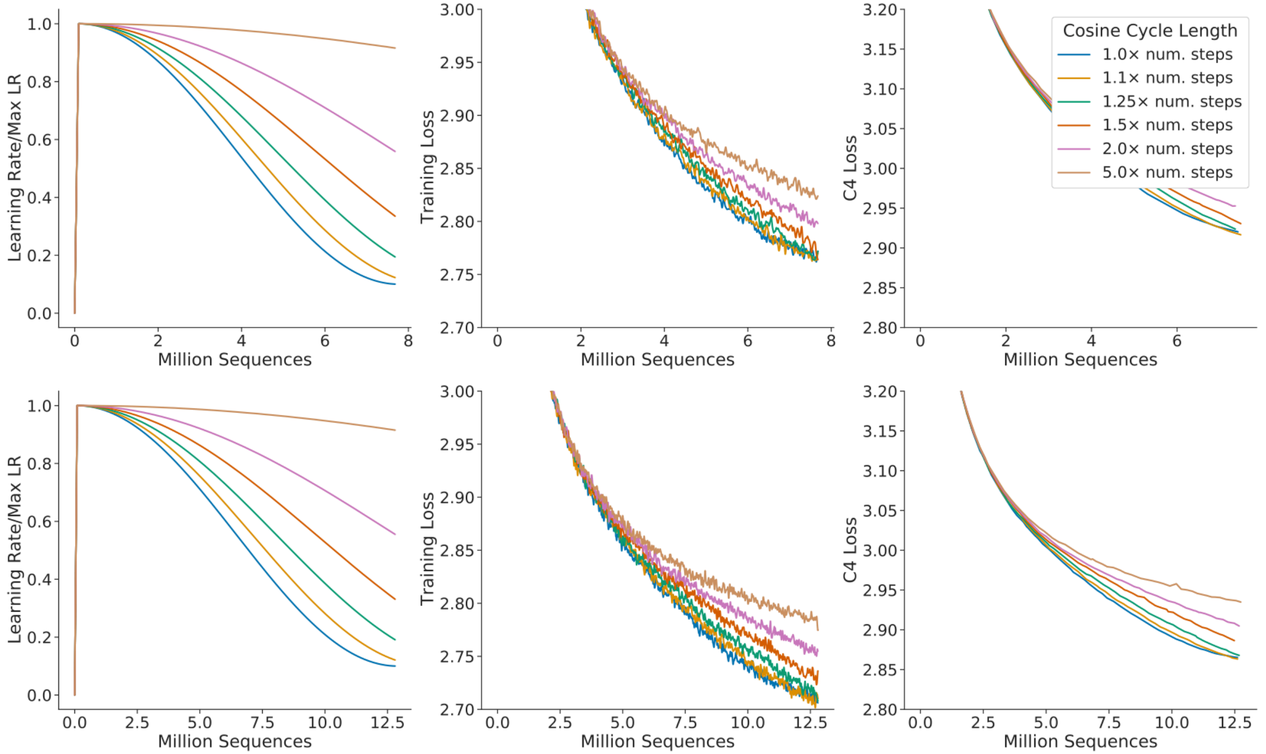
- Improperly configured cosine learning rate decay

- Recap: Cosine LR decay in PyTorch
  - `CosineAnnealingLR(optimizer, T_max, eta_min, last_epoch)`
  - `T_max: Maximum number of iterations`
  - `eta_min: Minimum learning rate. Default: 0.`



What Makes the Difference?

- Improperly configured cosine learning rate decay
- To get optimal loss, we should set **`T_max`** to actual end-training steps
  - But they used fixed scheduler so LR does not drop appropriately



[Hoffman+ 2022]

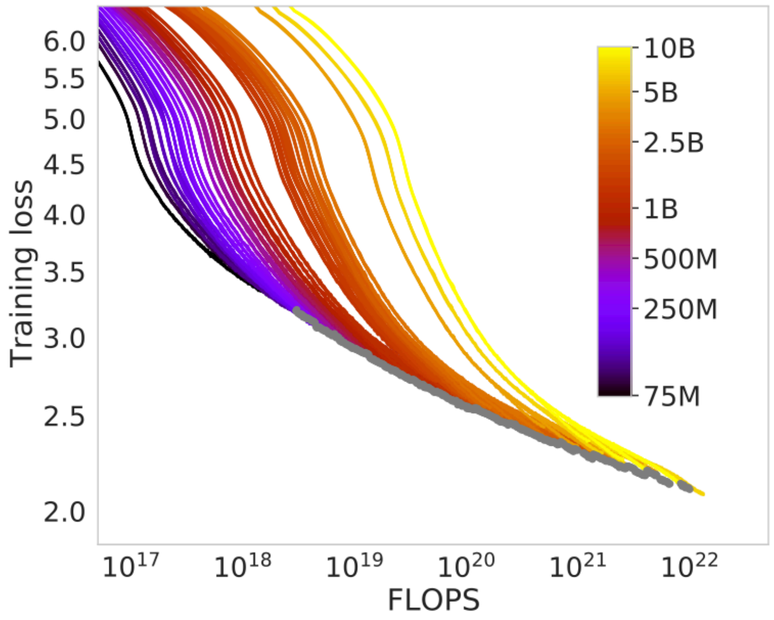
Chinchilla Scaling Laws [Hoffman+ 2022]

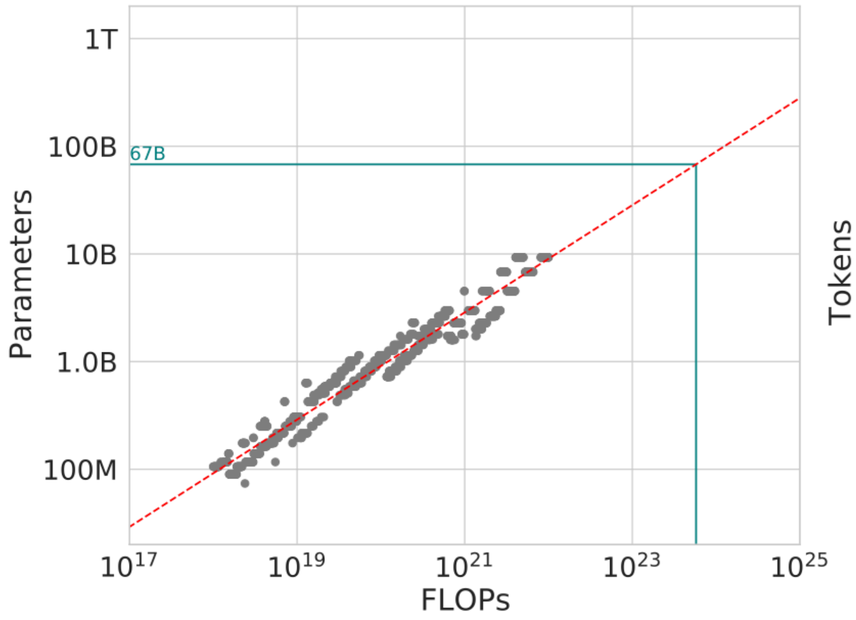
- They suggest 3 ways of fitting scaling laws
  - They mostly result similar constant

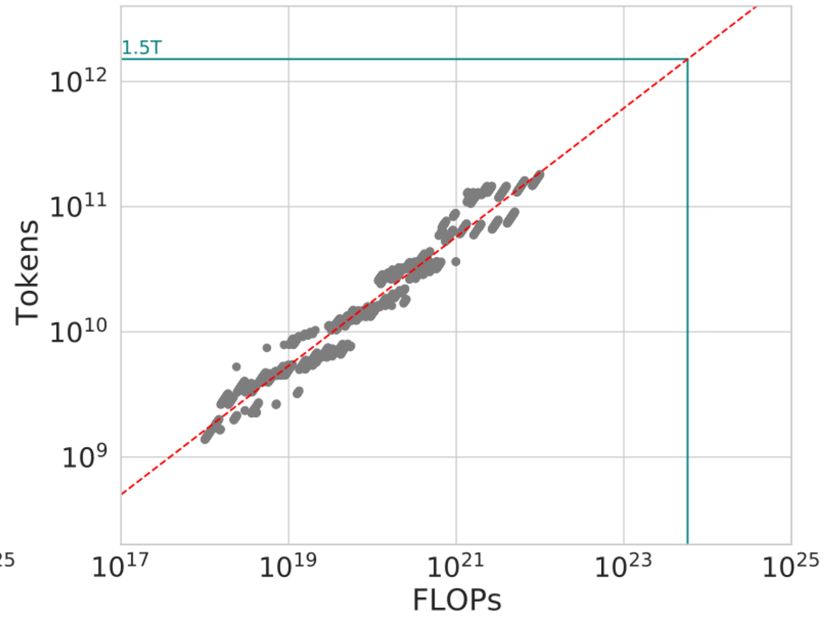
Approach	Coeff. a where $N_{opt} \propto C^a$	Coeff. b where $D_{opt} \propto C^b$
1. Minimum over training curves	0.50 (0.488, 0.502)	0.50 (0.501, 0.512)
2. IsoFLOP profiles	0.49 (0.462, 0.534)	0.51 (0.483, 0.529)
3. Parametric modelling of the loss	0.46 (0.454, 0.455)	0.54 (0.542, 0.543)
Kaplan et al. (2020)	(param) 0.73	(data) 0.27

Method 1: Minimum Over Runs [Hoffman+ 2022]

- Fix model sizes and vary number of training tokens
  - For each model size, train with four different training tokens
  - Then, get minimum loss envelopes and fit curve

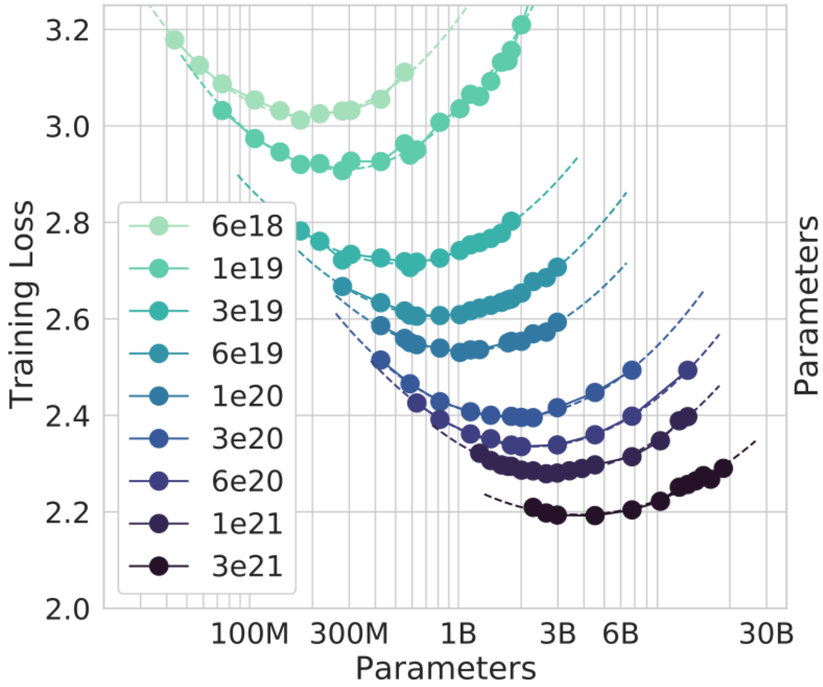


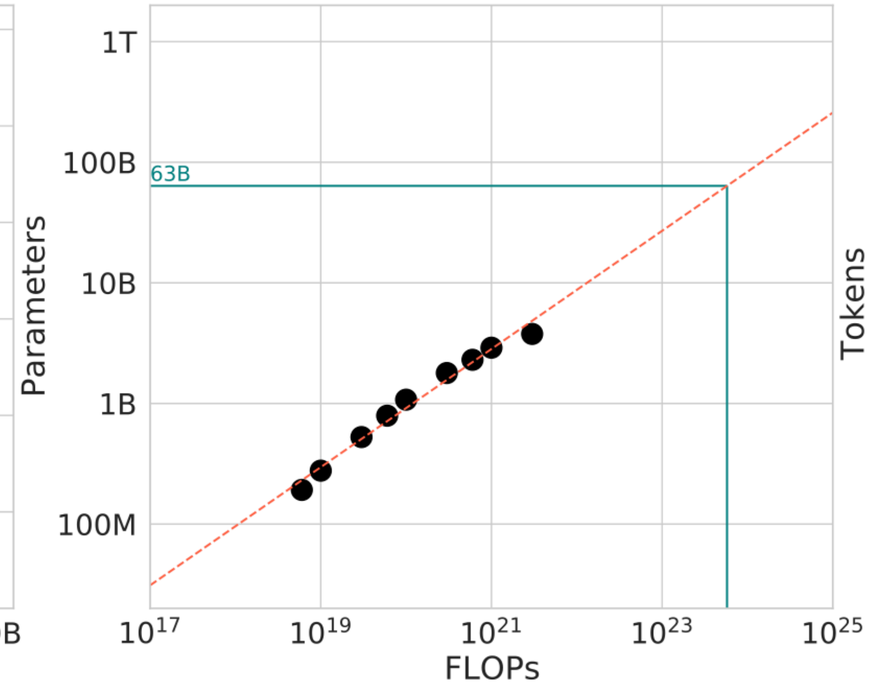


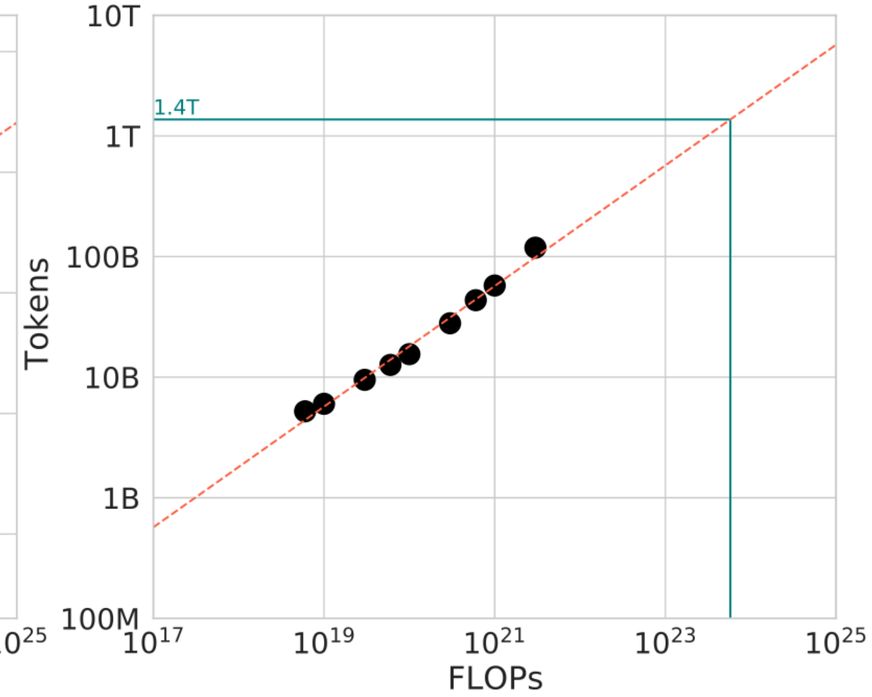


Method 2: IsoFLOPs [Hoffman+ 2022]

- Pick a range of FLOPs budgets, vary the model size for each FLOPs
- The min over these convex shapes form a power law

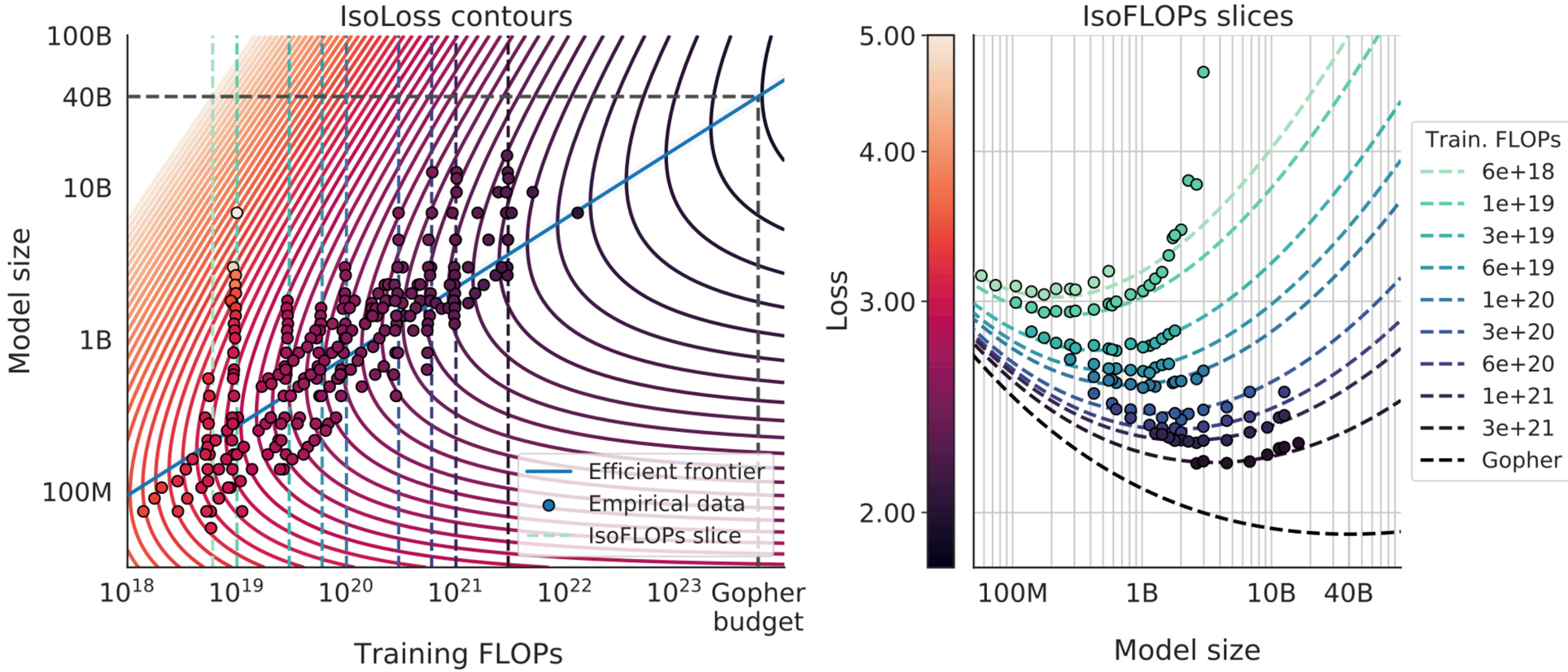






Method 3: Parametric Modeling [Hoffman+ 2022]

- Run a bunch of models on the size-data grid
- Use least squares to fit a joint scaling law



Chinchilla Model [Hoffman+ 2022]

Approach	Coeff. a where $N_{opt} \propto C^a$	Coeff. b where $D_{opt} \propto C^b$
1. Minimum over training curves	0.50 (0.488, 0.502)	0.50 (0.501, 0.512)
2. IsoFLOP profiles	0.49 (0.462, 0.534)	0.51 (0.483, 0.529)
3. Parametric modelling of the loss	0.46 (0.454, 0.455)	0.54 (0.542, 0.543)
Kaplan et al. (2020)	(param) 0.73	(data) 0.27

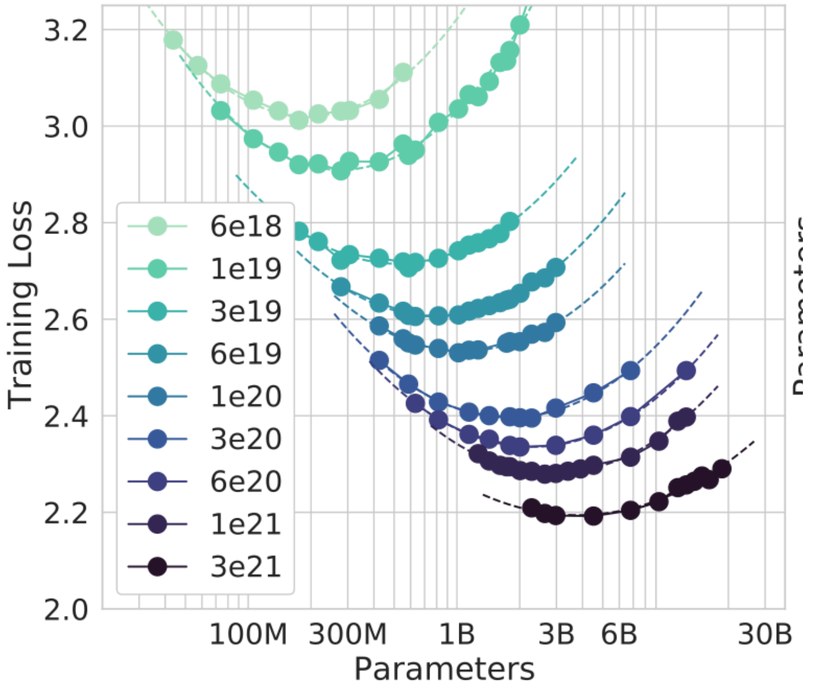
A new scaling laws and a new model

Model	Size (# Parameters)	Training Tokens
LaMDA (Thoppilan et al., 2022)	137 Billion	168 Billion
GPT-3 (Brown et al., 2020)	175 Billion	300 Billion
Jurassic (Lieber et al., 2021)	178 Billion	300 Billion
Gopher (Rae et al., 2021)	280 Billion	300 Billion
MT-NLG 530B (Smith et al., 2022)	530 Billion	270 Billion
Chinchilla	70 Billion	1.4 Trillion

But Train-Optimal May Not Be What We Want

- Chinchilla tells what gives the best model for fixed training compute
- But most of the compute in a real deployment is inference
  - So we should over-train

- Some examples:
  - GPT3 – 2 tokens / param
  - Chinchilla – 20 tokens / param
  - LLaMA 65B – 22 tokens / param
  - LLaMA 2 70B – 29 tokens / param
  - Mistral 7B – 110 tokens / param
  - LLaMA 3 70B – 215 tokens / param



Summary

- We can extend log-linearity (power law) to model size and compute

- Rethink the scenario
  - *Your friend has given you 10,000 H100s for a month, and asked you to build a good open source LLM*

- What do you do?
  - Lets us set the hyperparameters based on small models
  - Lets us make smart resource tradeoffs (model size vs. data size)
  - Scale up the model!

Summary

- Scaling laws are really surprising and useful tools

- Data scaling: understand how data affects models
- Model scaling: dramatically reduce costs for training
- Scaling as prediction: understand what problems can be brute forced

- Without scaling laws, we might…
  - Waste millions of dollars on "blind" trial-and-error
  - Produce "under-trained" models without knowing it
  - Fail to distinguish between "bad architecture" and "insufficient scale"
  - Lose the "extrapolation" advantage (train and pray 😢)