

Da målingen tømte en akse: enhetsmåling per beregningsfase, og et budsjett som ikke er innen rekkevidde

Ola Jacob Mjølven Iversen

Fellag AS (BoatNav), Kongsberg, Norge · olajacob@fellag.as

BoatNav Research Note 2026-04 · versjon 1.1, 1. august 2026 · preprint — ikke fagfellevurdert

Abstract. We report a pre-registered device pilot for maritime route planning on a physical handset, and a companion run of the identical code on a workstation. The interaction budget is missed by $25\text{--}32\times$ at p95, and the memory target by a factor of two; both outcomes fall on the bottom row of a decision table fixed before measurement, which makes the representation choice settled rather than open. Three findings go beyond the verdict. First, **the engine penalty is not one number**: measured per phase it ranges from $3.4\times$ to $48.8\times$, and the aggregate is not even well defined — two legitimate estimators disagree by more than a factor of two because the phase mix differs between machines. *Which phase looks like the bottleneck depends on the machine*: profiling on the workstation would have indicted post-simplification, the handset indicts grid construction. Second, the workstation — $14\times$ faster on plan time, no thermal ceiling — *also* misses the budget, which removes hardware as the explanation. Third, and most consequential for the architecture decision: the cost has two independent sources, and the pre-registered factor matrix contains **no viable level on one of its axes**. We also retract a mechanism claim from our own verification report: post-simplification is not demonstrably quadratic in raw vertex count, and the exponent cannot be established from these data.

Keywords: experimental algorithmics; mobile performance measurement; JavaScript engines; pre-registration; phase profiling; maritime route planning.

Sammendrag. Vi rapporterer en forhåndsregistrert enhetspilot for maritim ruteplanlegging på fysisk telefon, og en parallellkjøring av identisk kode på arbeidsstasjon. Interaksjonsbudsjettet bommes med $25\text{--}32\times$ ved p95 og minnemålet med en faktor to; begge lander på nederste rad i en beslutningstabell som ble låst før måling, hvilket gjør representasjonsvalget avgjort framfor åpent. Tre funn går ut over selve dommen. For det første er **motorstraffen ikke ett tall**: målt per fase spenner den fra $3,4\times$ til $48,8\times$, og aggregatet er ikke engang veldefinert — to legitime estimatorer spriker med over en faktor to fordi fasemiksen er ulik på de to maskinene. *Hvilken fase som ser ut som flaskehalsen, avhenger av maskinen*: profilering på arbeidsstasjonen ville pekt på etterforenklingen, telefonen peker på rutenettkonstruksjonen. For det andre bommer arbeidsstasjonen — $14\times$ raskere på plantid, uten termisk tak — *også* på budsjettet, hvilket fjerner maskinvaren som forklaring. For det tredje, og viktigst for arkitekturvalget: kostnaden har to uavhengige kilder, og den forhåndsregistrerte faktormatrisen har **ingen brukbar verdi på én av sine akser**. Vi trekker samtidig en mekanismepåstand fra vår egen verifikasjonsrapport: etterforenklingen er ikke påviselig kvadratisk i råstiens verteksantall, og eksponenten lar seg ikke fastslå av disse dataene.

1 Innledning

Et ruteslag i BoatNav er et råd. Skipperen skal lese ruta, forstå hvorfor knekkpunktene ligger der de ligger, justere den for hånd og styre etter den. Det stiller et interaksjonskrav som er lett å undervurdere: forslaget må komme raskt nok til at det oppleves som et svar og ikke som en jobb maskinen setter i gang.

Research Note 2026-01 [1] tallfestet kravet som *løsningsmål M5*: p95 under ett sekund for turer kortere enn 5 nautiske mil, og under tre sekunder opp til 30 nm — målt på *mål-enheten*, i release-bygg, på Hermes-motoren. Alle tidstall i den rapporten var derimot målt i Node på arbeidsstasjon, og notatet førte selv opp konsekvensen under trusler mot validitet: forholdet mellom de to måtte etableres i en pilot før hovedforsøket.

Dette notatet innfrir det løftet, og rapporterer utfallet. Piloten er forhåndsregistrert i sin helhet — instansutvalg, kjøreplan, gyldighetskontroller og, avgjørende, en *beslutningstabell* som oversetter måleresultatet til en konsekvens for fase 2. Tabellen ble skrevet før første kjøring, slik at kriteriet ikke kan velges etter at resultatet er kjent.

1.1 Bidrag

1. En forhåndsregistrert enhetsmåling med 210 kjøringar på fysisk telefon, uten feil, med rene gyldighetskontroller for termisk drift og rekkefølgeeffekt (§2–§3).
2. **Node↔Hermes-forholdet målt per fase**, med det metodiske hovedfunnet at et samlet forholdstall verken er informativt eller veldefinert (§4).
3. Påvisning av at *flaskehalsen bytter fase med maskinen*, og at et arkitekturvalg tatt på profilering fra feil motor ville angrepet feil fase (§5).
4. Identifisering av en **tom akse** i den forhåndsregistrerte faktormatrisen: for gitterfamilien finnes ingen registrert etterbehandlingsverdi som oppfyller M3, M4 og M5 samtidig (§6).
5. En tilbaketrekking av vår egen mekanismepåstand, med begrunnelsen for hvorfor to punkter ikke kan bære en eksponent (§7).

Tabell 1: Protokollen slik den ble kjørt. Alle krav i forhåndsregistreringen er innfridd.

Krav	Målt
Bygg	cce197a, release, iOS 26.5
Oppvarmingsregel	per-instans — den registrerte
Kjøringer	210 (150 målte, 60 oppvarming)
Feil	0
Batteri	80 % → 35 %, aldri under gulvet
Lader	ikke tilkoblet i noen kjøring
Strømsparing	av i alle 210

2 Apparatet

Piloten kjørte på en iPhone 14 (iOS 26.5) i release-bygg med git-stempel cce197a, mot det frosne flis-snapshotet fra fase 0 (manifest 9f45be95, 3 167 fliser). Utvalget er 30 instanser trukket fra det frosne korpuset på 273 (5cd635aa), stratifisert over farvannstype og lengdebånd.

Protokollen er motbalansert: fem runder, rekkefølgen snus annenhver runde, og to oppvarmingskjøringer foran hver instans i første runde. Det gir 210 kjøringer, hvorav 150 er måledata.

2.1 Gyldighet

Motbalanseringen finnes for å avdekke termisk drift over en økt på nesten to timer, og kontrollen er ren. Samme instans i runde 5 mot runde 1 gir median **1,003** (p10 0,949, p90 1,021) over de 30 instansene. Baklengs mot forlengs runder gir median **1,010**, med retningen delt. Ingen av delene er en reell effekt.

Termisk tilstand var *ikke tilgjengelig* fra plattformen i noen kjøring — iOS eksponerer ProcessInfo.thermalState, men verken React Native 0.86 eller Expo SDK 57 gir den videre til JavaScript. Feltet er ført som *ikke-tilgjengelig* framfor utelatt. Vi kan derfor ikke utelukke struping direkte, bare indirekte gjennom fraværet av drift.

2.2 Node-siden

Etter telefonen, og bare etter, ble identisk kode kjørt i Node 20 på arbeidsstasjon mot samme 30 instanser og samme flis-snapshot, med $n = 20$ — 600 kjøringer. Rekkefølgen er registrert og ikke tilfeldig: med Node-tallene først er det menneskelig å justere telefonoppsettet til forholdet ser rimelig ut, og et forhold man har justert seg fram til måler oppsettet framfor maskinene.

En port kjøres før forholdet regnes: grunn, verteksFor og verteksEtter må være identiske i alle 30 instanser. De er det. De to motorene regner altså samme rute, og forholdet måler motorene framfor to ulike beregninger.

3 M5: budsjettet bommes med 25–32 ganger

Responstiden måles som msPlan — konstruksjon, søk, forenkling og verifisering — med flishenting holdt utenfor, som registrert.

Tabell 2: M5 på målenheten. Begge båndene med budsjett lander på nederste rad i den forhåndsregistrerte beslutningstabellen.

Bånd	n	p50	p95	p95/budsj.
< 5 nm	55	14,2 s	31,6 s	31,6×
5–30 nm	50	12,3 s	76,2 s	25,4×
> 30 nm	45	17,3 s	159,2 s	—

Tabell 3: Forholdet telefon/Node per fase. *Median* er median over instansenes egne medianer; *pooled* er snitt over alle kjøringer. n er antall instanser medianen hviler på etter kvantiseringsfilteret.

Fase	Median	p10	p90	Pooled
Konstruksjon	17,1	11,8	19,6	17,5
Søk	48,8	47,1	68,2	51,4
Forenkling	3,7	3,6	3,9	3,5
Verifisering	3,4	3,4	3,7	3,2
Plan (sum)	14,0	5,5	18,0	6,3

Beslutningstabellen i arbeidspakkens §4 har fire rader. Den nederste lyder:

> 5× — **budsjettet er ikke oppnåelig med uniformt rutenett**. Fase 2 skal da behandle representasjonsvalget som avgjort av dette, ikke som en åpen matrise.

Begge bånd lander der, med fem til seks ganger klaring. Ingen av de tre mildere radene er i nærheten. Slutningen følger av en regel skrevet før måling, ikke av en vurdering i etterkant.

M6 — minnetaket på 150 MB — bommes tilsvarende: 76 av 150 kjøringer ligger over, med median 154,3 MB og maksimum 407,8 MB. Og tallet er en *nedre* grense: Hermes fører ingen høyvannsmerke-teller, så prøver tatt på fasegrenser kan ha gått glipp av høyere topper mellom dem.

4 Forholdet er ikke ett tall

Løftet fra 2026-01 var å etablere Node↔Hermes-forholdet. Det viktigste resultatet er at spørsmålet slik det er stilt ikke har ett svar.

Fasene ligger mellom 3,4× og 48,8× — en faktor 14 mellom ytterpunktene. Et samlet forholdstall ville derfor tatt feil med opptil en størrelsesorden for enkeltfaser.

Det samlede tallet er heller ikke veldefinert. Per fase er de to estimatorene enige (17,1/17,5; 48,8/51,4; 3,7/3,5; 3,4/3,2). For summen spriker de — 14,0 mot 6,3. Grunnen er at aggregatet arver fasemiksen framfor å måle motoren, og fasemiksen er ulik på de to maskinene. Et prosjekt som oppgir «Hermes er k ganger tregere enn Node» oppgir dermed en egenskap ved sin egen arbeidsmengde, ikke ved motorene.

4.1 Klokka setter grensen for hva som kan sammenliknes

Fasetimeren i implementasjonen bruker Date.now() — 1 ms oppløsning, og ikke-monoton. På telefonen er det uvesentlig (konstruksjon 9 003 ms). På Node er det ikke: medi-

Tabell 4: Andel av snittlig plantid per motor. Samme kode, samme innndata.

Fase	Telefon	Node
Konstruksjon	53 %	19 %
Forenkling	45 %	80 %
Søk	2 %	0 %

Tabell 5: M5-tallene målt på Node. Beslutningstabellen gjelder telefonen og anvendes ikke her — men forklaringen «telefonen er bare treg» er utelukket.

Bånd	p50	p95	p95/budsj.
< 5 nm	1,0 s	4,4 s	4,4×
5–30 nm	1,0 s	15,0 s	5,0×
> 30 nm	1,0 s	45,2 s	—

an søketid er **3 ms**, og 140 av 600 kjøringene måler 0 ms.

Forholdene for søk og verifisering er derfor regnet over bare 6 av 30 instanser — de der Node-medianen er ≥ 20 ms, slik at kvantiseringsfeilen holder seg under 2,5 %. Tallene er robuste (median 48,8 mot pooled 51,4), men de hviler på et smalt utvalg og skal leses slik.

Ikke-monotonien ga ingen skade i denne økta: null negative fasetider i 210 kjøringene, og msPlan avviker maksimalt 2 ms fra summen av fasene. Et klokkehopp ville brutt begge. Men det er en egenskap ved *dataene*, ikke en garanti fra instrumentet, og timeren skal over på en monoton klokke med sub-ms oppløsning før fase 2 måler søket.

5 Flaskehalsen bytter fase med maskinen

Fordi motorstraffen er ulik per fase, er fasemiksen ulik på de to maskinene — og rangeringen snur.

Profilering på Node ville pekt på etterforenklingen. Telefonen peker på rutenettkonstruksjonen ved medianen. Begge profilene er korrekte for sin maskin. Hermes straffer konstruksjonen $17\times$ og forenklingen bare $3,7\times$, så hvilken som dominerer avhenger av hvilken maskin man måler på.

Konsekvensen er en metoderegulering vi registrerer: **en faseprofil er ikke en egenskap ved koden alene**, og skal alltid oppgis med maskin. Et arkitekturvalg tatt på en profil målt på annen motor enn målenheten kan angripe feil flaskehals.

5.1 Maskinvaren er ikke forklaringen

Node er en arbeidsstasjon med JIT, uten termiske grenser og uten batteri, og den er 14 ganger raskere enn telefonen på plantid.

Medianen ligger på 1,0 s mot et krav om 1 s, og p95 er fire til fem ganger over. **Et uniformt rutenett når ikke budsjettet på en maskin som er 14 ganger raskere enn målenheten.** Det er et sterkere utsagn enn telefonmålingen alene, fordi det fjerner maskinvaren fra forklaringsrommet.

Tabell 6: Fasene på målenheten. Forenklingen har lav median og ekstrem hale.

Fase	Median	p95	Maks
Konstruksjon	9 003 ms	29 426 ms	31 187 ms
Søk	132 ms	2 683 ms	3 827 ms
Forenkling	664 ms	50 842 ms	151 432 ms
Verifisering	24 ms	99 ms	134 ms

6 Kostnaden har to kilder — og matrisen har en tom akse

Ved medianen eier konstruksjonen **91,3 %** av plantiden og søket 0,9 %. Det bekrefter 2026-01s hovedfunn skarpere enn tidligere målt, og det gjør det på tvers av en motor som straffer nettopp søket hardest ($48,8\times$). Fasen med størst motorstraff er fortsatt uten betydning.

Men i halen stemmer bildet ikke. Etterforenklingen har median 664 ms og maksimum **151 sekunder** — lengre enn konstruksjonens maksimum. I 25 av 150 kjøringene overstiger den ti sekunder, og i 17 % av kjøringene er den større enn konstruksjonen.

Det betyr at å fikse konstruksjonen alene ikke løser M5. Selv med et gratis rutenett ville den verste instansen brukt 151 sekunder. De to kostnadene må angripes hver for seg.

6.1 Hullet

Den forhåndsregistrerte faktormatrisen har fire akser. Aksen for etterbehandling har nøyaktig to nivåer: **P0** (av) og **P1** (grådig eksakt etterforenkling, dagens).

For gitterfamilien er begge utelukket:

- **P0 ryker på M3.** Løsningsmålets egen begrunnelse i forskningsopplegget lyder «utelukk rå 8-naboskapsgrid uten etterbehandling (8,2 % verstefall)» mot et krav om $\varepsilon \leq 5$ %. P0 ryker også på M4: 62 rutenett-punkter i ett forslag var utgangspunktet for kravet om høyst 15 vertekser.
- **P1 ryker på M5**, med opptil 151 sekunder.

For diskretiseringene D1–D3 finnes det altså ingen registrert verdi på P-aksen som oppfyller M3, M4 og M5 samtidig. Matrisen er ikke bare smalnet av beslutningsregelen — den er tom på én akse.

Det er dette notatets viktigste konsekvens for arkitekturvalget, og det er et funn måleoppsettet ikke var designet for å gjøre. Utveien er enten nye P-nivåer, eller en diskretisering utenfor gitterfamilien.

6.2 To argumenter konvergerer

Synlighetsgraf (D4) og klaringsfelt med GVD-skjelett (D5) gir få vertekser og nær-eksakte stier *uten* etterforenkling. Piloten skyver dermed mot dem av to uavhengige grunner: gitteret er dyrt å bygge, og gitteret trenger en etterbehandling alternativene i prinsippet klarer seg uten. To uavhengige argumenter som peker samme vei er sterkere enn ett — men

merk at synlighetsgrafens egen svakhet (kollaps i tett geometri) er kjent fra litteraturen [1] og ikke målt av oss ennå.

7 En mekanismepåstand trekkes

Verifikasjonsrapporten for denne piloten sluttet i sin første utgave at etterforenklingen er *kvadratisk* i råstiens verteksantall. Slutningen hvilte på én sammenlikning mellom to instanser: 612 vertekser mot 323, der kvadratisk skalering forutsier 3,59 og målt forhold var 3,26.

To punkter kan ikke skille $v^{1,5}$ fra v^2 fra v^2 ganget med noe annet. Node-siden gir 20 repetisjoner per instans uten termisk støy, og lar eksponenten estimeres over alle 16 instansene med nok signal:

$$ms = 0,22 \cdot v^{1,59}, \quad R^2 = 0,384.$$

Forholdet ms/v^2 spenner **65 ganger** mellom instansene, og to instanser med praktisk talt likt verteksantall — 314 og 337 — skiller **42 ganger** i tid. Verteksantallet forklarer 38 % av variasjonen. De to opprinnelig valgte instansene hadde tilfeldigvis liknende geometri.

Den forventede formen er v^2 ganget med *geometritetthet langs stien*, og den er forenlig med dataene: forenklingen tester hvert kandidatsnarveg eksakt mot hindringsgeometrien, så kostnaden per test avhenger av hvor mange kanter segmentet møter. Åpen kyst møter få, indre skjærgård mange. Men antall geometritester logges ikke, så modellen kan ikke prøves.

Det som står uten forbehold: forenklingen *vokser* med verteksantallet (log-log-korrelasjon 0,62), den eier halen, og maksimum er 151 sekunder. Eksponenten er ukjent.

7.1 Hvorfor dette ble funnet

Verken denne retraksjonen eller den andre rettelsen i rapporten — to ulike medianbegreper begge kalt «median» — ble funnet ved gjennomlesning. **Begge ble funnet fordi en uavhengig andre måling ikke stemte med den første.** Node-siden ble regnet med en nyskrevet analysefunksjon, tallene stemte ikke, og uenigheten tvang fram en kontroll.

En kilde til et tall hadde ikke avdekket noe av det. Det er samme mønster som i 2026-03, der ni ubrukte referanser overlevde flere gjennomlesninger og ble oppdaget av en mekanisk port.

8 Hva som ikke er målt

- **Antall geometritester** i forenklingen logges ikke. Kostnadsmodellen kan derfor ikke prøves, og eksponenten forblir uavklart.
- **Termisk tilstand** var ikke tilgjengelig. Struping kan bare utelukkes indirekte.
- **Heap på Node** er ikke sammenliknet. V8 og Hermes har ulike samlere og ulike objektrepresentasjoner; M6 er en telefonstørrelse.
- **Korpusets dekning av trangt farvann** er svak, og det er der både M5-bommen og perturbasjonsfølsomhe-

ten [2] konsentrerer seg. Trusselen er registrert; fase 2 skal rapportere per stratum framfor som aggregat.

9 Anbefalinger

1. **Representasjonsvalget behandles som avgjort**, etter den forhåndsregistrerte regelen. Ingenting i dette notatet gjenåpner det, og et dårlig resultat skal ikke i seg selv gi flere runder.
2. **P-aksen må få nye nivåer før matrisen kjøres.** Slik den står, har den ingen brukbar verdi for gitterfamilien.
3. **Instrumenter forenklingen først.** Uten antall geometritester velges P-nivåer i blinde.
4. **Sett en gjennomførbarhetsport foran fase 2.** Forbehandling forutsetter et verdensankret gitter [2] og en verdensankret oppløsningsstige innenfor bytedusjettet for offline kartpakker. Får den ikke plass, er hele familier utilgjengelige, og da skal de ikke måles.
5. **Oppgi alltid maskin med en faseprofil**, og ta aldri et arkitekturvalg på profilering gjort på annen motor enn målenheten.

Tilgjengelighet

Alle artefakter er frosset med SHA-256: korpus b13a98ec, fasit 96020cd3, baseline 7acd5b15, flismanifest 9f45be95, fase 1-utvalget 5cd635aa. Pilotloggen (210 kjøring) og Node-kjøringen (600 kjøring) foreligger som artefakter med egne hasher. Analysen er et kjørbart skript framfor en beskrivelse, slik at forholdstallene kan reproduceres uten forfatteren. Sjøkartdata er Kartverkets, under NLOD.

Erklæringer

Interessekonflikt: Forfatteren eier Fellag AS, som utvikler BoatNav. Artikkelens måler Fellags eget produkt, og rapporterer at produktet bommer på sine egne mål.

KI-assistanse: Arbeidet er utført med gjennomgående assistanse fra Claude (Anthropic) — i analysen, i verifiseringen og i manuskriptet. Piloten er kjørt av implementasjonsspor og verifisert av et forskningsspor med egen analysefunksjon. Retraksjonen i §7 og konvensjonsrettelsen den følger av er verifikasjonssporets funn, rettet mot dets egen tidligere rapport. Alle måletall stammer fra BoatNav-implementasjonen og de frosne artefaktene. Forfatteren står ansvarlig for innholdet.

Bidragsroller (CRediT): Conceptualization, Methodology, Software, Investigation, Formal analysis, Validation, Visualization og Writing er utført av forfatteren med KI-assistanse som beskrevet over. *Validation* er holdt uavhengig av implementeringen: analysen er reimplementert framfor gjenbrukt, og uenigheten mellom de to implementasjonene er det som avdekket begge rettelsene.

Data og kode: Pilotloggen, Node-kjøringen og analyseskriptet er kommittet i prosjektets repositorier. Måleharnessen planlegges åpnet sammen med resultatartikkelen; inntil da tilgjengelig på forespørsel.

Ansvarsavgrensning: BoatNav gir veiledende rutforslag for fritidsbruk og er ikke et godkjent navigasjonshjelpemiddel.

Versjon 1.1 (1. august 2026): tre formuleringer skrevet om etter at en leser påpekte at «bortforklares» tilskrives et motiv framfor å beskrive hva forhåndsregistrering gjør. Den fjerner frihetsgrader, og virker uavhengig av hensikt — det retoriske ordet gjorde påstanden snevrere enn den er. Innført som mekanisk port B8 i fagfelleprotokollen. Ingen tall eller konklusjoner er endret.

Referanser

- [1] O. J. M. Iversen. Automatisk ruteplanlegging for fritidsbåter i norsk skjærgård: problemkarakterisering, målinger og forskningsagenda. *BoatNav Research Note 2026-01*, Fellag AS, 2026.
- [2] O. J. M. Iversen. Rutenettet er ankret til spør­ringen, ikke til verden: start­punktsensitivitet i mobil ruteplanlegging for skjærgård. *BoatNav Research Note 2026-02*, Fellag AS, 2026.