Accelerating Google's Flood Forecasting Initiative with Tensor Processing Units

Damien Pierce
dmpierce@google.com

Sept 2020

with V. Anisimov, A. Boral, Y. Chen, L. Hu, S. Nevo, Y. Shafi
Google Research

Flood forecasting

- Affects hundreds of millions of people
- Thousands of fatalities per year
- Flood forecasting is an effective mitigation tool
    - Can reduce fatalities and economic impacts by a third*

*J. Malilay: Floods. In *The Public Health Consequences of Disasters*, Oxford University Press, 1997

Flood forecasting ingredients

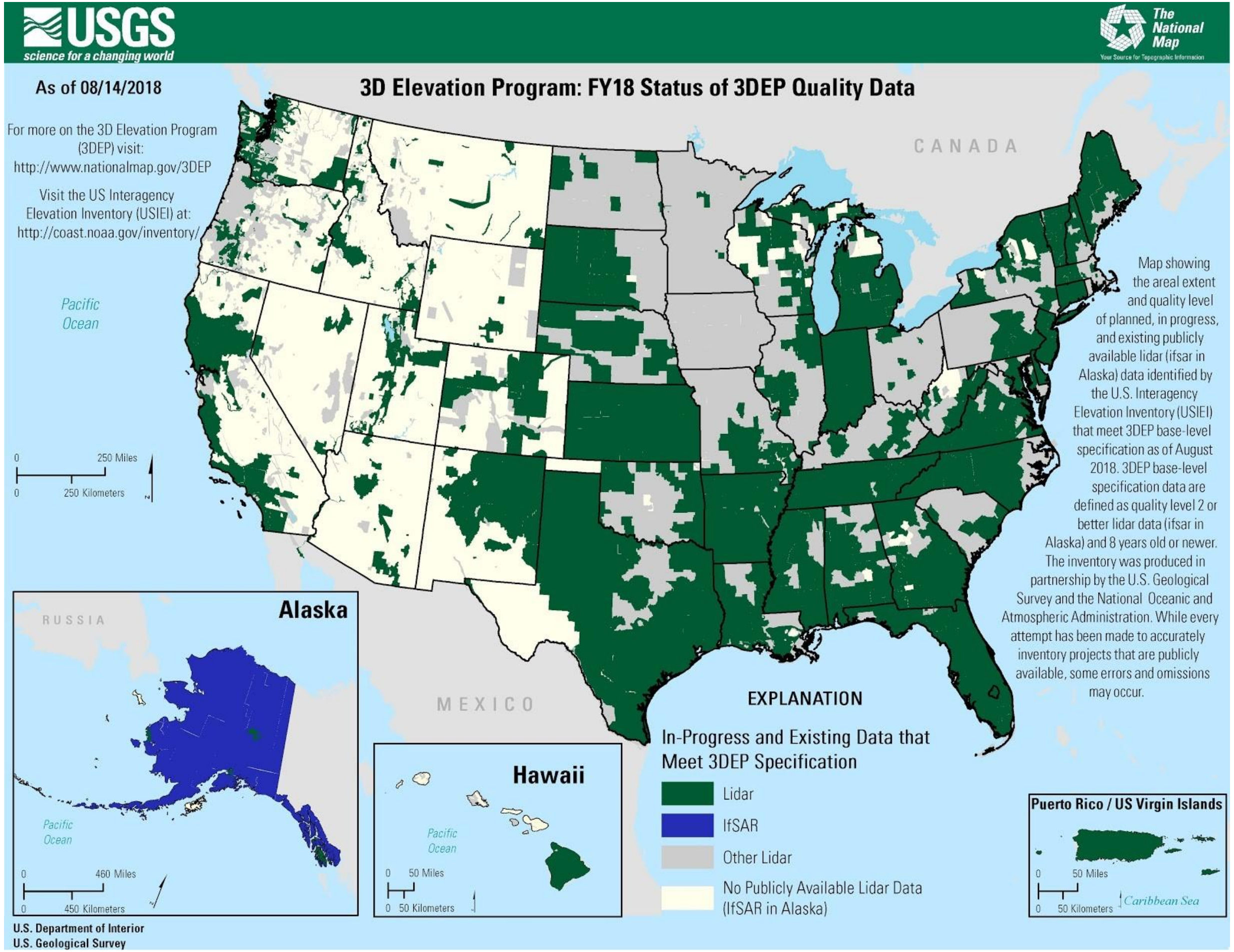
- Real-time and forecasted water level measurements
- High resolution Digital Elevation Models (DEMs)
- Forecasting techniques: some combination of
  - Hydrological modeling
  - **Hydraulic modeling**
  - Machine learning

Hydraulic modeling: 2D shallow water equations

$$\frac{\partial h}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} = 0$$

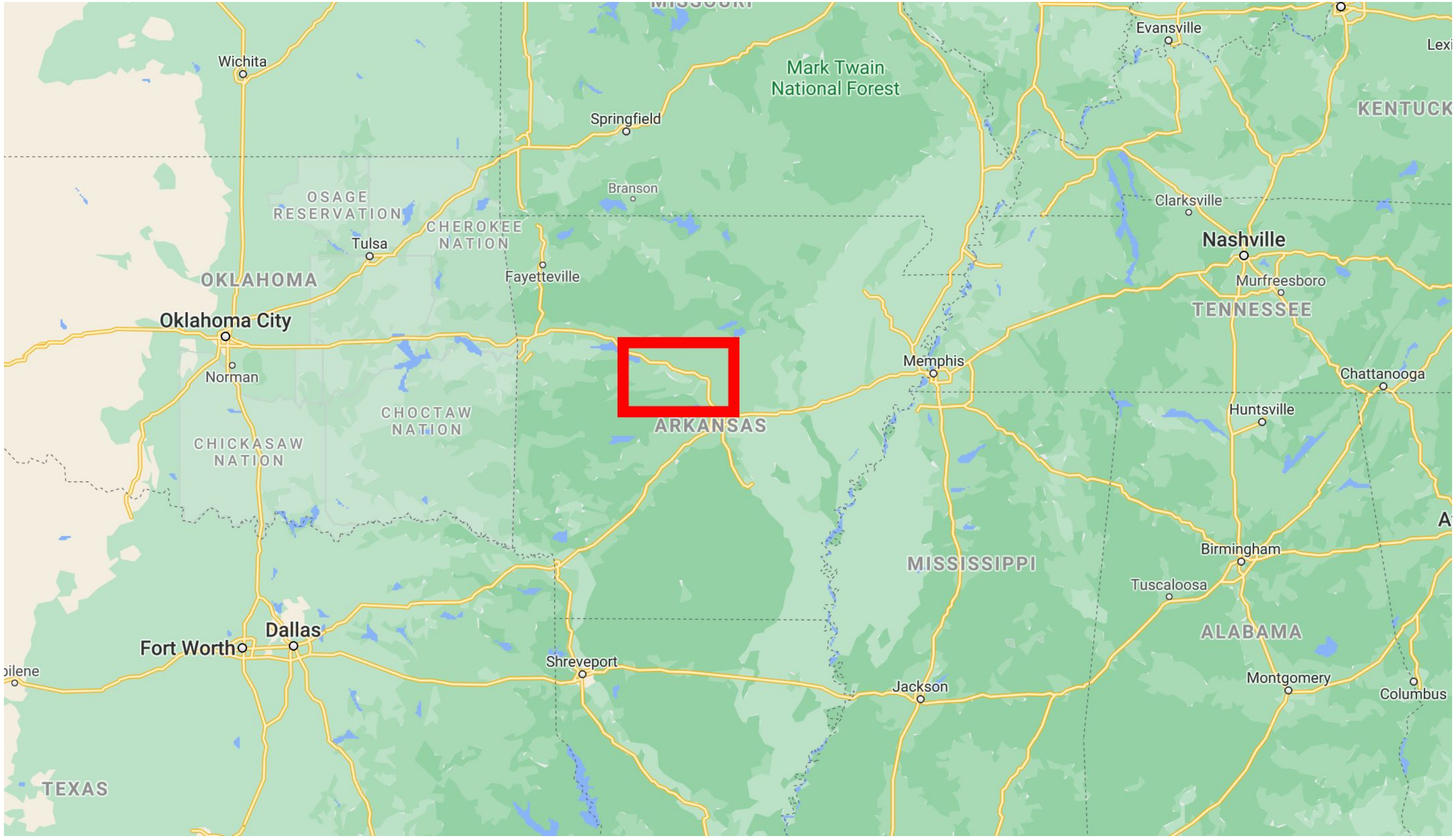
$$\frac{\partial q_i}{\partial t} + gh\frac{\partial (h + z)}{\partial i} + \frac{gn^2}{h^{7/3}}\|\mathbf{q}\|q_i = 0, \quad i \in \{x, y\}$$

q = flux [discharge per unit width, L^2 / T]
h = water height
z = surface elevation
n = Manning friction coefficient

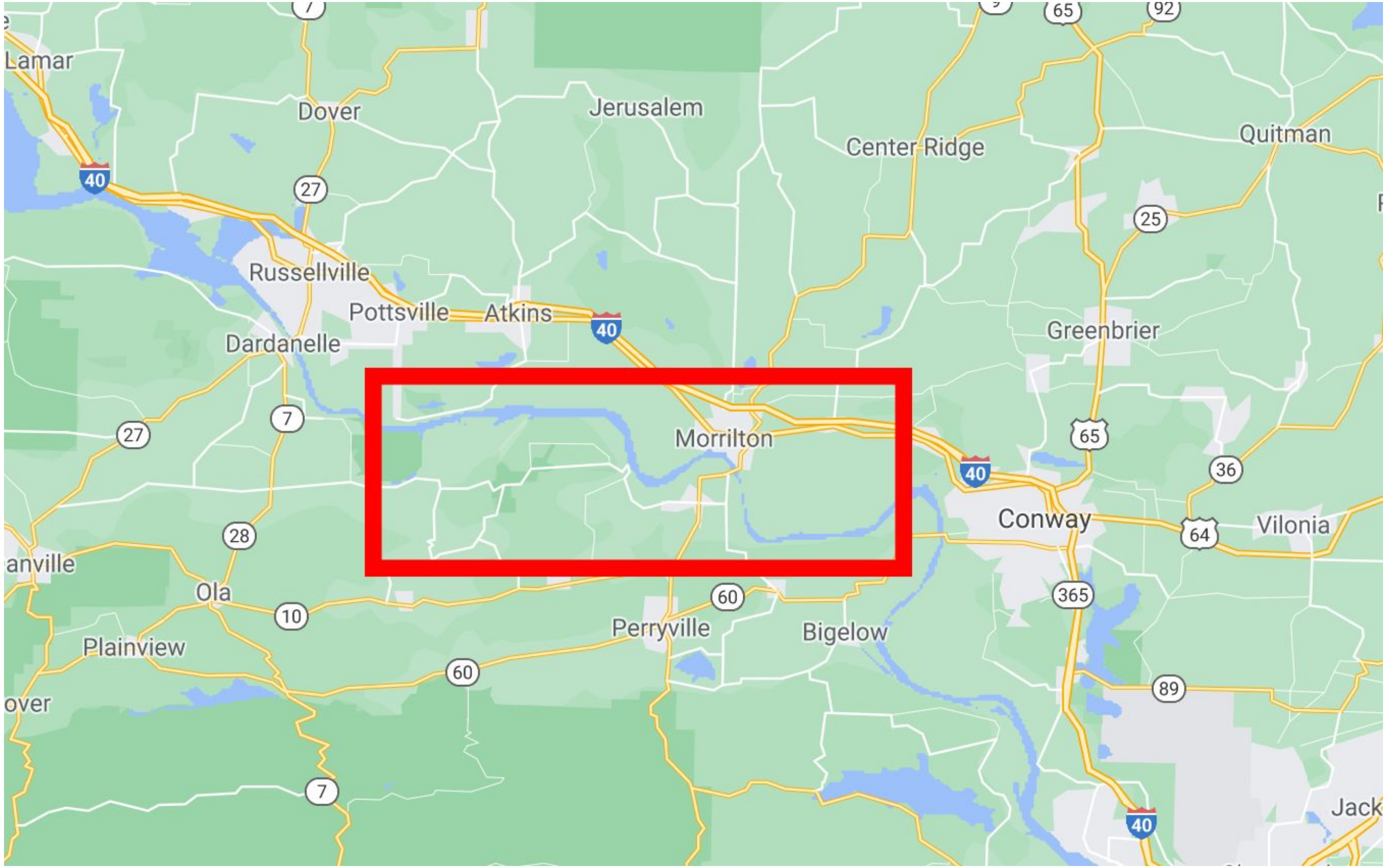


USGS 3DEP Map As of Aug 2018

https://nationalmap.gov/preview/3DEP


Wichita
Springfield
Mark Twain
National Forest
Evansville
KENTUCK
Branson
OSAGE
RESERVATION
CHEROKEE
NATION
Tulsa
Clarksville
Nashville
Fayetteville
OKLAHOMA
Murfreesboro
TENNESSEE
Oklahoma City
Memphis
Norman
Chattanooga
Huntsville
CHOCTAW
NATION
ARKANSAS
CHICKASAW
NATION
MISSISSIPPI
Birmingham
Tuscaloosa
Dallas
Fort Worth
ALABAMA
Shreveport
Jackson
Montgomery
Columbus
TEXAS

Zooming in...



Arkansas River

Flooded in May, 2019

Region modeled:

990 sq km
244k acres

aspect ratio:
2.15 : 1

Almost 1 billion square meters → 1 billion grid points in 1m simulation

Hydraulic Model Simulation

- Main parameter is the discharge at the input boundary (volume of water per unit time)
- Run to (close to) steady state (2 days)
- Run with various discharges
- Results compared to satellite images
- → Discharge = 15k m^3/s

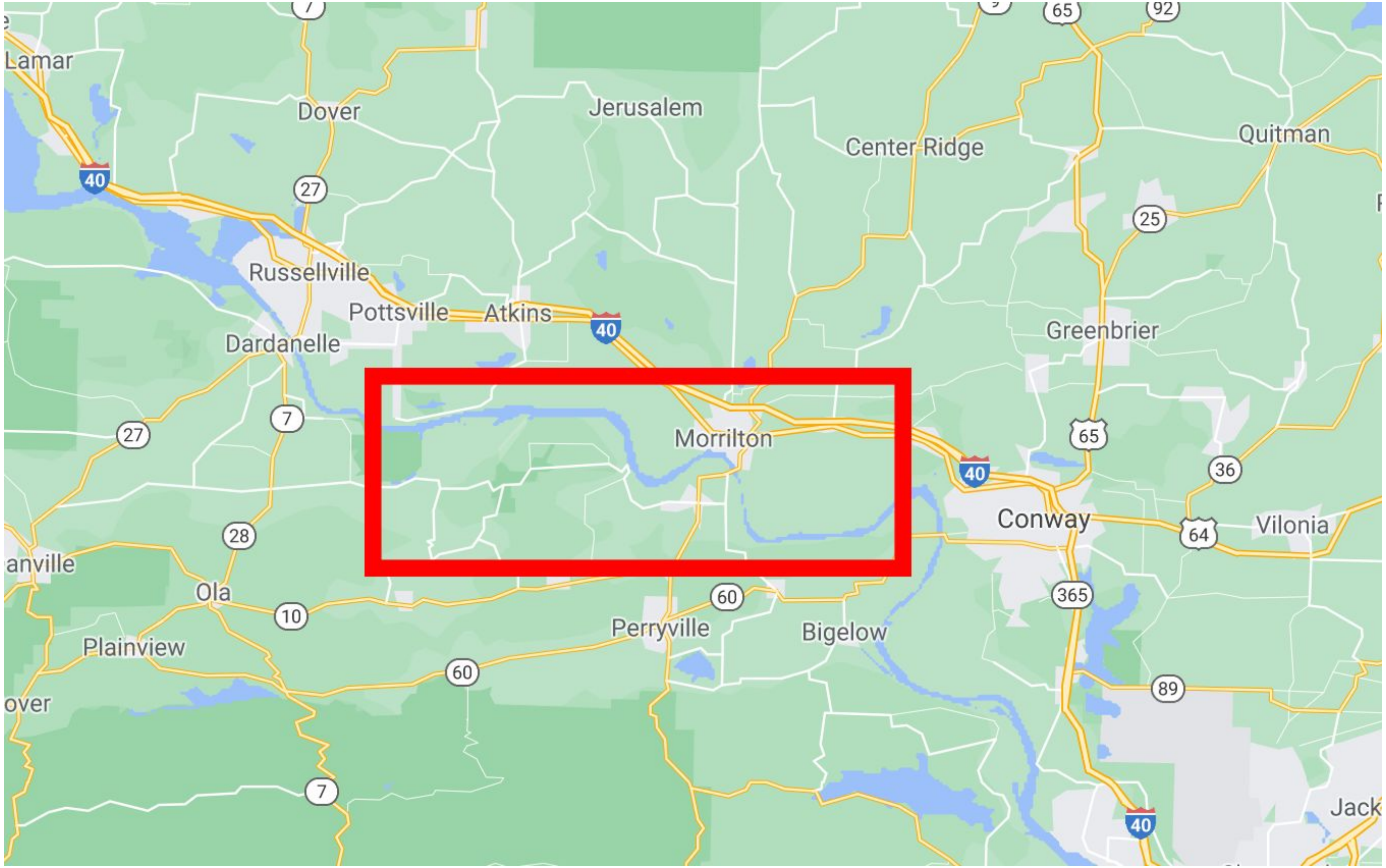
Google Research

Table 3. Streamgage information related to calculation of annual exceedance probability for the May to June 2019 flood event along Arkansas.

[AEP, annual exceedance probability; USGS, U.S. Geological Survey; ft, foot; ft^3/s, cubic foot per second]

USGS streamgage number[1]	USGS streamgage name	Date of peak streamflow	Peak streamflow for May to June 2019 flood			
			Peak gage height (ft)	Peak streamflow (ft^3/s)	Rank of peak streamflow in record	Number of annual peaks
07152500	Arkansas River at Ralston, Okla.	5/23/2019	22.14	185,000	1	43
07164500	Arkansas River at Tulsa, Okla.	5/29/2019	23.51	277,000	2	55
07165570	Arkansas River near Haskell, Okla.	5/29/2019	24.24	286,000	1	47
07194500	Arkansas River near Muskogee, Okla.[2]	5/26/2019	46.39	600,000	1	33
07250550	Arkansas River at James W. Trimble L&D near Van Buren, Ark.[3]	5/31/2019	406.96	570,000	1	50
07258000	Arkansas River at Dardanelle, Ark.[3]	5/30/2019	45.91	565,000	1	50
07263450	Arkansas River at Murray Dam near Little Rock, Ark.[3]	6/4/2019	259.75	520,000	1	50

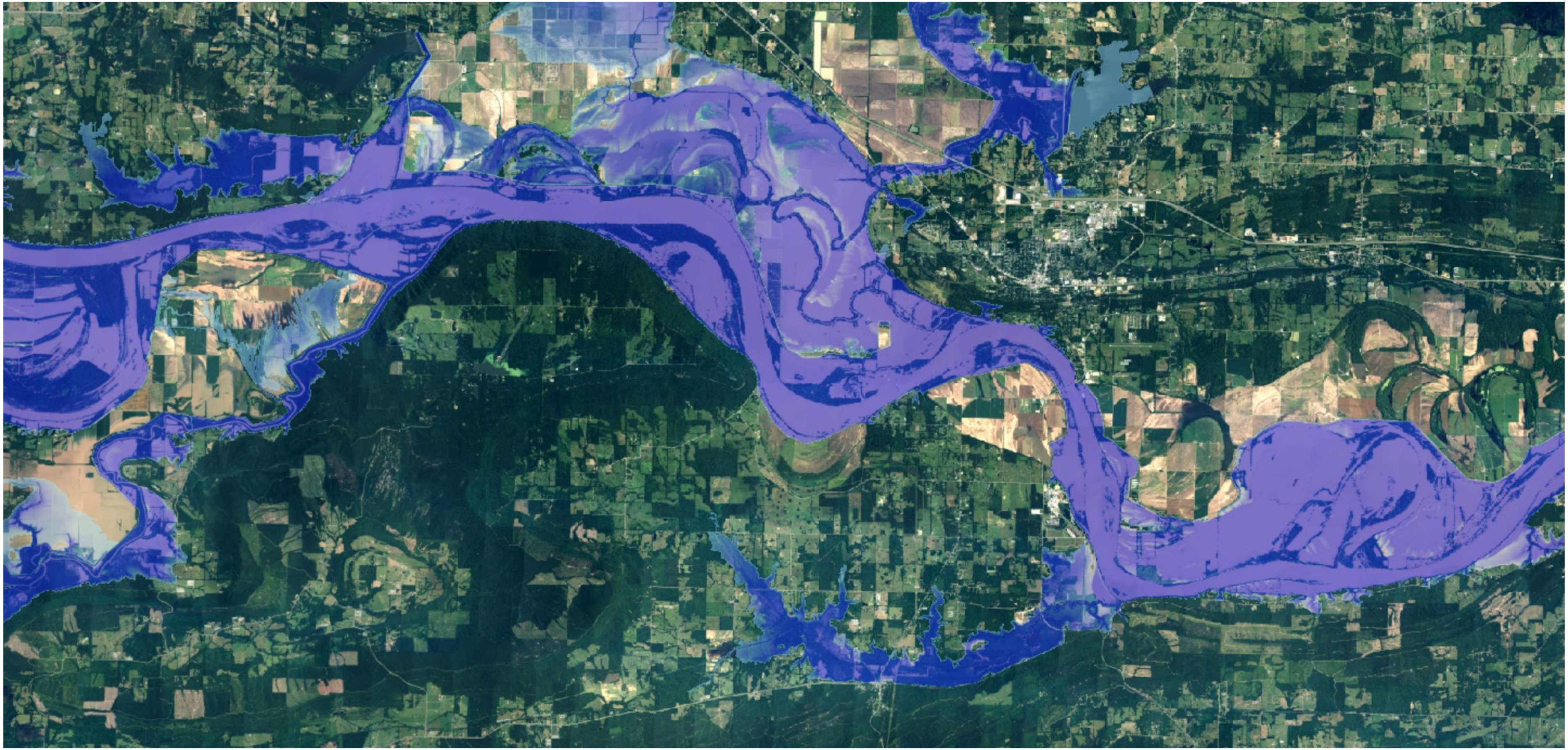
$565k\ ft^3 / s = 16k\ m^3 / s$


Lamar
Dover
Jerusalem
Center Ridge
Quitman
Russellville
Pottsville
Atkins
Dardanelle
Greenbrier
Morrilton
Conway
Vilonia
anville
Ola
Plainview
Perryville
Bigelow
over
Jack
40
27
7
28
10
60
65
92
25
36
64
365
89

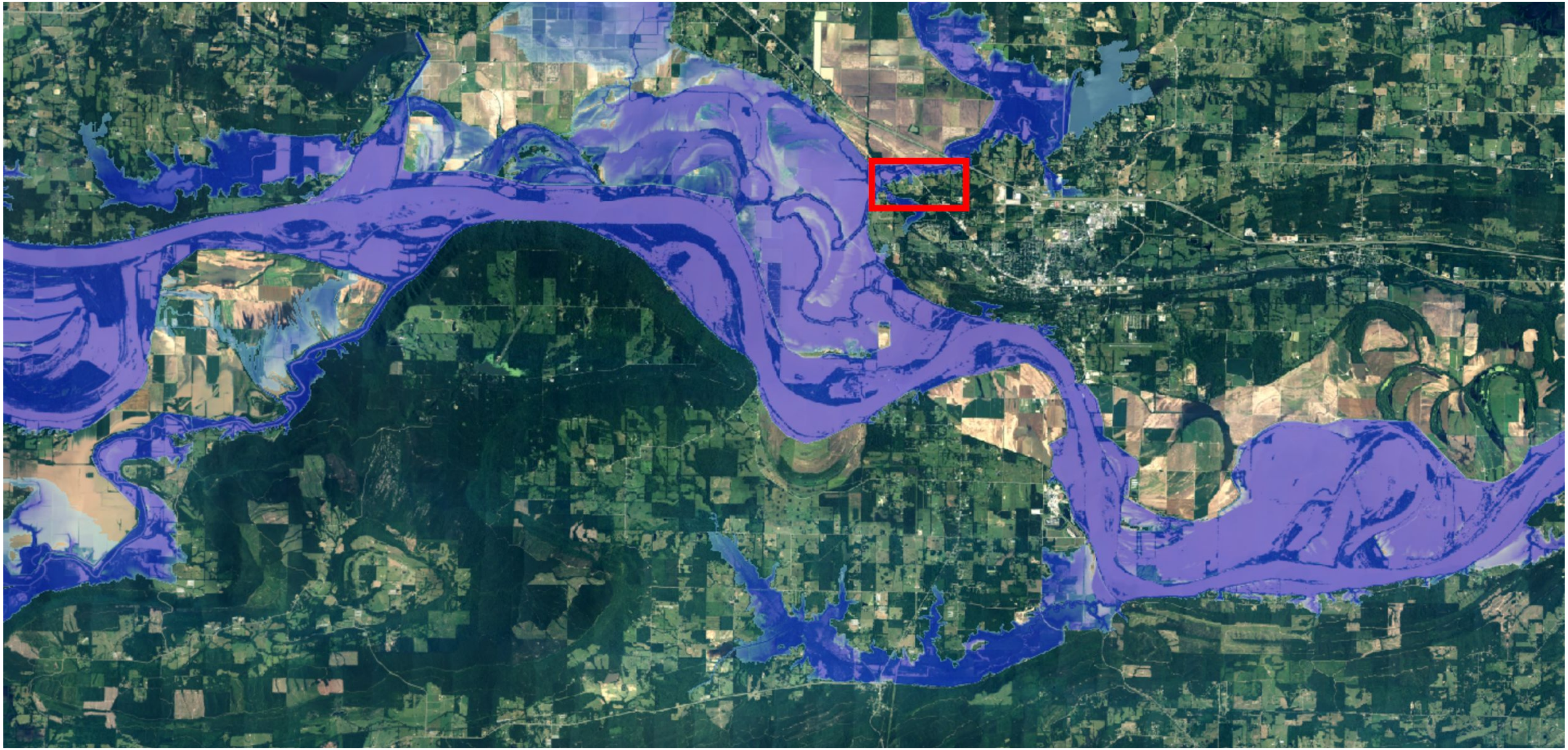
Zooming in...

satellite image from 5/19/2019

https://gis.arkansas.gov



15k discharge simulation overlay



Zooming in...

Observations about sim result

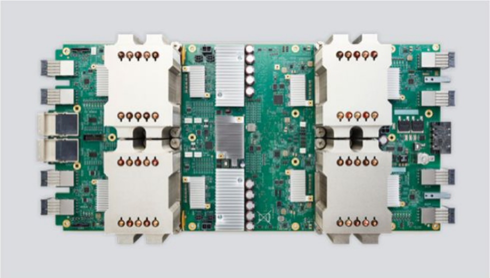
- USGS 3DEP Lidar provides an excellent DEM
  - captures bare earth beneath trees
  - includes bathymetry
- Running simulation on 64 CPU cores can take typically O(days)
  - How to speed this up? Days → minutes?

Hardware accelerators

GPUs are well equipped to train AI models

- Thousands of cores
- Large memory bandwidth
- Matrix multiplication

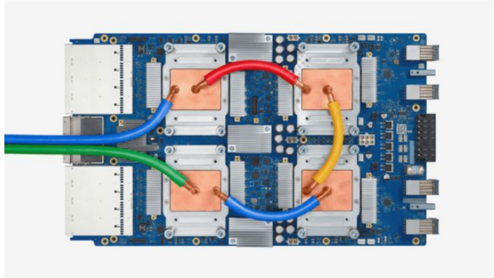
Since 2016 Google has launched **TPUs** specifically to increase AI performance → **Also great for HPC**



Cloud TPU v2

180 teraflops

64 GB High Bandwidth Memory (HBM)



Cloud TPU v3

420 teraflops

128 GB HBM

1 Cloud TPU has

4 chips

2 cores/chip

8 cores

Cloud TPU v2 Pod

11.5 petaflops

4 TB HBM

2-D toroidal mesh network

Cloud TPU v3 Pod

100+ petaflops

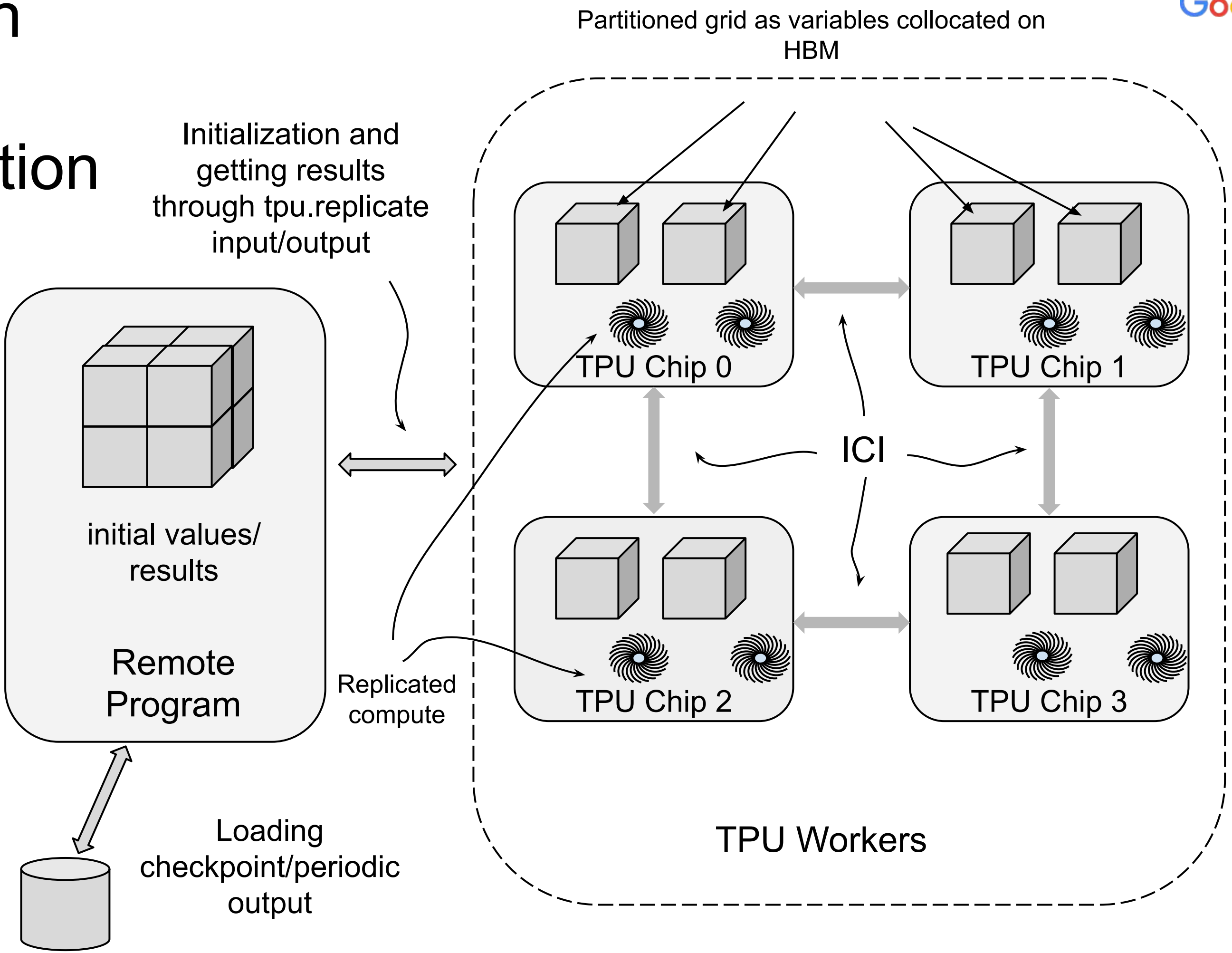
32 TB HBM

2-D toroidal mesh network

256 Cloud TPUs form a v3 Pod

2048 cores

Distribution and Parallelization



Google Research

Simulation performance comparison

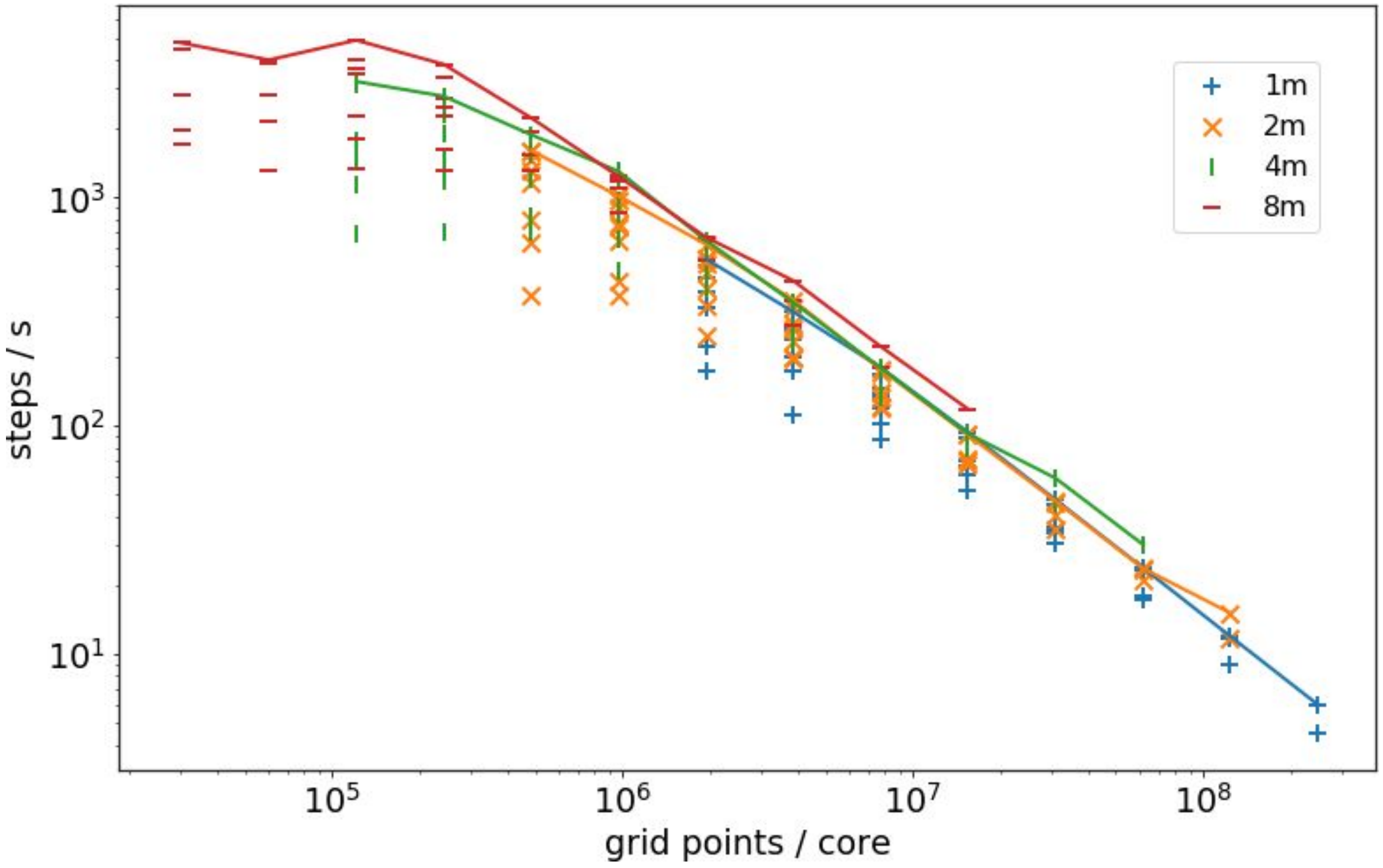
Single CPU core vs. single TPU core

Intel Xeon E5-16504 v4 @ 3.6 GHz vs. Google Cloud TPU v3

CPU-TPU Comparison Results				
Cell Size (m)	Grid Points (M)	CPU steps/s	TPU steps/s	Speed Up
4	62	0.26	30.22	118
8	15	1.04	118	114

4m resolution for 1.7 million steps:
77 days for 1 CPU core vs.
16 hours for 1 TPU core → 512 cores → **9 minutes**

Arkansas Flood Simulation Performance for 1 to 512 cores



Layout of TPU Cores

- 2D simulation → 32 cores (e.g.) can have various assignments per axis: 1x32, 2x16, 4x8, etc.
- In many HPC settings, a more square per-core grid will be most efficient (8x4 in this case since the grid aspect ratio is ~2)
- TPUs have very high bandwidth, so latency dominates
  - The most extreme layouts (e.g. 32x1 or 1x32) are most efficient in this 2D case

Time to compute 1.7 million steps				
Resolution	**8 cores**	**32 cores**	**128 cores**	**512 cores**
8m	43 mins	13 mins	5.9 mins	6.1 mins
4m	2.7 hours	44 min	15 mins	8.9 mins
2m	10 hours	2.7 hours	46 mins	18 mins
1m	40 hours	10 hours	2.7 hours	53 mins

1.728 million steps = 2 simulation days if `dt = 0.1` sec.

Google Research

Weak Scaling Efficiencies				
Resolution	**8 cores**	**32 cores**	**128 cores**	**512 cores**
8m	43 mins	13 mins	5.9 mins	6.1 mins
4m	2.7 hours	44 mins 97%	15 mins 84%	8.9 mins 66%
2m	10 hours	2.7 hours 99%	46 mins 93%	18 mins 72%
1m	40 hours	10 hours 100%	2.7 hours 100%	53 mins 80%

Strong Scaling Efficiencies				
Resolution	**8 cores**	**32 cores**	**128 cores**	**512 cores**
8m	43 mins	13 mins 83%	5.9 mins 46%	6.1 mins 11%
4m	2.7 hours	44 mins 91%	15 mins 66%	8.9 mins 28%
2m	10 hours	2.7 hours 94%	46 mins 83%	18 mins 54%
1m	40 hours	10 hours 99%	2.7 hours 94%	53 mins 70%

Flood forecasting using hydraulic models

- In **steady-state** rivers, many simulations with different discharges are typically done offline, before flood season.
- During flooding, given actual and predicted stream gauge measurements, the correct discharge is picked out and alerts are sent out.
- Changing run times from days to minutes allows for a real-time approach.
- Also, real time approaches are needed in case of dynamic rivers (**non-steady-state**).

Conclusion

- Hydraulic flood simulations are a useful tool in flood forecasting
- Running simulations on TPUs can dramatically decrease run times
  - Scaling results shown for Arkansas flood simulation
- Running on a fleet of TPUs opens the possibility for real time approaches in both steady-state and dynamic river cases (e.g. variational data assimilation)
- AI: TPUs can readily generate data sets for machine learning training
- Paper in progress; GCP Python interactive notebook with flood simulation will be made available