



SPE 173246

Accelerating Tight Reservoir Workflows With GPUs

K. Mukundakrishnan, K. Esler, D. Dembeck, V. Natoli, J. Shumway, Y. Zhang, Stone Ridge Technology, and J. R. Gilman, H. Meng, iReservoir.com, Inc.

Copyright 2015, Society of Petroleum Engineers

This paper was prepared for presentation at the SPE Reservoir Simulation Symposium held in Houston, Texas, USA, 23-25 February 2015.

This paper was selected for presentation by an SPE program committee following review of information contained in an abstract submitted by the author(s). Contents of the paper have not been reviewed by the Society of Petroleum Engineers and are subject to correction by the author(s). The material does not necessarily reflect any position of the Society of Petroleum Engineers, its officers, or members. Electronic reproduction, distribution, or storage of any part of this paper without the written consent of the Society of Petroleum Engineers is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of SPE copyright.

Abstract

There are numerous complex characteristics that impact the long-term decline behavior of wells in tight oil and gas reservoirs. Numerical simulation with fine spatial discretization is required to capture important characteristics such as: the presence of natural fractures (stimulated and unstimulated); changes in total fluid mobility and compressibility; heterogeneous matrix and fracture properties (including permeability loss due to compaction); early flow transients; and nonuniform stimulation during hydraulic fracturing. This discretization requirement, when combined with the need to include multiple wells to capture interference effects, can result in model sizes in the tens of millions of cells over a few 640-acre sections. While these model sizes can sometimes be addressed with current-generation simulators, the excessive run time limits the ability to simulate multiple realizations for history matching and sensitivity analysis. In practice, lower fidelity simulations are often substituted.

We present our efforts to help eliminate this tradeoff by building a fully-implicit black-oil simulator that combines recent advances in simulation algorithms with the high performance of GPUs. All major computational tasks are executed on GPU, including property evaluation, Jacobian construction and assembly, and linear solution with CPR-AMG. This approach allows models with many millions of cells to be simulated within minutes on a single workstation with multiple GPUs. For example, on a tiled SPE 10 model with 55 million cells and 250 wells, we simulate 2000 days of production in 20 minutes using eight GPUs. We summarize our approach to address the challenges in building a fine-grained, scalable simulator. We discuss two challenges in particular: the need to expose massive parallelism while retaining the robustness of linear solvers; and managing the complexity of the many features required by engineers for practical application.

We discuss how we apply this fully-accelerated technology to increase fidelity and throughput on tight reservoir workflows, improving our understanding of the complex nature of production decline and possible long-term well interference. Furthermore, we illustrate workflows for simulations based on detailed reservoir/fracture descriptions.

Introduction

Development of low permeability tight oil and tight gas reservoirs in North America accelerated in recent years due to the development and large-scale deployment of high-volume hydraulic stimulation technology. Extracting maximum value from these reservoirs require a means to determine optimal values for development and production parameters which minimize cost without sacrificing long-term recovery. These parameters include the interval between fractured stages, the ideal horizontal and vertical spacing between wells, and operating constraints for production such as the bottom-hole pressure. Furthermore, reliable prediction of long-term recovery is essential to assess the economic viability of individual plays, especially as fluctuating oil prices dip toward the break-even points in some formations.

Several simplistic analytical solutions exist for history matching and production forecasting in tight oil reservoirs. An approximate solution to the dual-porosity model that is valid during the transient and late time flow from the fracture and matrix in unconventional oil reservoirs has been presented in Ogunyomi et al. (2014). A five-region model that is a generalization of both the trilinear model of (Brown et al., 2011) and the enhanced frac region model of (Stalgorova and Mattar, 2012b) has been proposed in Stalgorova and Mattar (2012a). However, these models include oversimplified assumptions such as single-phase flow, constant or slight compressibility, negligible rock compaction, and idealized fracture geometries. Many engineers also rely on empirically-fit decline characterizations for production forecasting (Okouma et al., 2012). Many of these decline characterization methods were originally developed for conventional reservoirs in the middle part of the last century (Arps, 1945). For many unconventional plays, quality historical production data is only available for a short period. Because processes such as the evolution of free gas, changes in the total compressibility and mobility, and possible well interference come into effect only later in production, it is unclear whether this data is sufficient to calibrate declines with reliable long-term predictive value (Gilman and Fadaei, 2013).

As with conventional plays, increasingly detailed reservoir models are now available for unconventional through the incorporation of data derived from seismic surveys, well logs, core images, and fluid property measurements. Detailed numerical simulations of such models that addresses both the geologic and dynamic flow characteristics can offer greater predictive accuracy for the long-term well performance. Lin et al. (2014) carried out simulations of geomodel-based tight oil reservoirs for different fracture geometries and their effects on oil production. They concluded that simple planar fracture geometries overestimated oil recoveries. Grid sizes trying to honor geologic complexity such as stacked pay intervals, and completion complexity such as long well length, numerous hydraulic stages, and closely spaced wells can run into millions to tens-of-millions of cells. Such models take from several hours to days to run with standard simulation technology. When combined with a rapid drilling cadence, the multiple realizations required along with their run times can make workflows based on numerical simulations impractical, hence the retreat to simple decline characterizations, oversimplified analytical models, or simplistic single-well numerical models.

However, recent developments in both computational hardware performance and the refinement of solver algorithms can provide the ability to simulate large complex models in much less time. In particular, graphical processing units (GPUs) provide an extremely dense computational platform with high memory bandwidth and arithmetic throughput. The benefits of GPUs have already been realized in seismic imaging and implementation has matured beyond research stage to large scale production use (Foltinek et al., 2009). Typical imaging algorithms, such as reverse time migration, require the porting and optimization of only a few key computational kernels. However, extracting a similar acceleration from GPUs for reservoir simulation requires a more extensive effort due to the complexity and abundance of kernels needed to support a wide range of model features. The challenges involved in a building a fully accelerated simulator have been described in detail in Esler et al. (2014) and will be briefly outlined below.

Description of the GPU simulator

The simulator we developed is based on a finite-volume discretization of the black-oil model. It employs standard approaches for the description of fluid, rock fluid and rock properties, and incorporates common well models and control mechanisms. For fractured reservoirs, including unconventional, it supports both dual-porosity and dual-permeability models, in addition to reversible and irreversible rock compaction with anisotropic transmissibility modification as a function of overburden pressure. To ensure robust convergence, the coupled, nonlinear system of equations is solved in a fully-implicit manner using a Newton-Raphson iterative method to achieve quadratic convergence. The resulting linear systems are solved iteratively with Krylov methods and preconditioners with very strong convergence properties. Notable among these is a constrained pressure residual method with algebraic multigrid pressure preconditioning (CPR-AMG), which has become a work-horse of many modern reservoir simulators.

Implementing a simulator on the GPUs must address a number of key challenges. These include selecting and optimizing robust solver algorithms, avoiding bottlenecks due to unaccelerated code paths, exposing sufficient parallelism, and carefully managing the limited amount of memory available on a GPU card. Earlier efforts to address some of these challenges have focused primarily on accelerating solvers and preconditioners since they constitute the bulk of the computation times (Appleyard et al., 2011, Yu et al., 2012, Tchelepi and Zhou, 2013, Bayat and Killough, 2013, Chen et al., 2013). The optimal choice of solver algorithm depends on the characteristics of the underlying differential equations, the size and complexity of the model and the hardware resources available. The pressure and saturation subsystems exhibit contrasting behaviors: the pressure subsystem is elliptic, while saturation is mostly hyperbolic. Preconditioners such as Jacobi and some incomplete factorizations, are well-suited for solving the hyperbolic subsystem because of their localized nature of error reduction. For elliptic pressure subsystems with strong heterogeneity, an algebraic multigrid (AMG) preconditioner is found to be the best performer for the large problem sizes solved in this study. Considering all the above constraints, we have chosen a highly optimized two-stage, constrained pressure residual (CPR)-AMG preconditioner for our simulations (Esler et al., 2014).

The performance gains that are achievable through the acceleration of the solvers and preconditioners alone are restricted by Amdahl's law. In many cases, the linear solver may comprise less than 70% of the total run time, and the remaining 30% of unaccelerated code can limit performance. We avoid these bottlenecks by implementing all major computation on GPU, including property evaluation, Jacobian assembly and the solution of the linear equations. The additional benefit of addressing Amdahl's law is that the large and expensive CPU-to-GPU transfers are avoided.

Attaining high performance for each of these tasks on GPUs requires exposing thousands of simultaneous, independent threads of work. As such, a naïve translation of algorithms designed for CPU will not typically result in performance gains. Some computational tasks, such as the evaluation of cell-based properties have abundant inherent parallelism. Implementing them on the GPU is therefore relatively straight-forward. However, many robust preconditioners were originally formulated as sequential algorithms, and a significant reformulation is required to expose parallelism. This is particularly true of AMG, which has complex data dependencies and attaining significant speedup is a non-trivial task (Esler et al., 2012).

Fully-implicit simulations are dominated by sparse matrix operations, the performance of which is typically limited by memory bandwidth rather than floating-point throughput. Hence, it is critical to adopt efficient data layouts which maximize the utilization of available bandwidth for all computations. As an example, strided memory access arising from an array-of-structures (AoS) data layout typically used in legacy codes results in very poor GPU performance and thus, has been completely avoided. Furthermore, we adopt matrix formats that have been optimized for the characteristics of the GPU memory hierarchy.

Another important challenge in developing a GPU simulator is managing the limited total memory on the hardware. In our simulator design, only the data necessary for computational kernel execution is stored on GPU. Frequently-used data—including reservoir state vectors, property tables, the Jacobian matrix, and the residual vectors are always kept resident in GPU memory. Essential data exchanges with the CPUs and with other GPUs are often overlapped with simultaneous computation, which can mitigate or eliminate the communication overhead.

By incorporating all of the above strategies, we have been able to simulate two simultaneous realizations of a 16 million cell dual-porosity / dual-permeability tight oil unconventional reservoir model on a single 2U server (**Fig. 1**) for 20 years of production with very fast run times. In this paper, we explore the application of our massively parallel GPU simulator in accelerating multiple realization workflows of a real-field unconventional tight oil reservoir play.

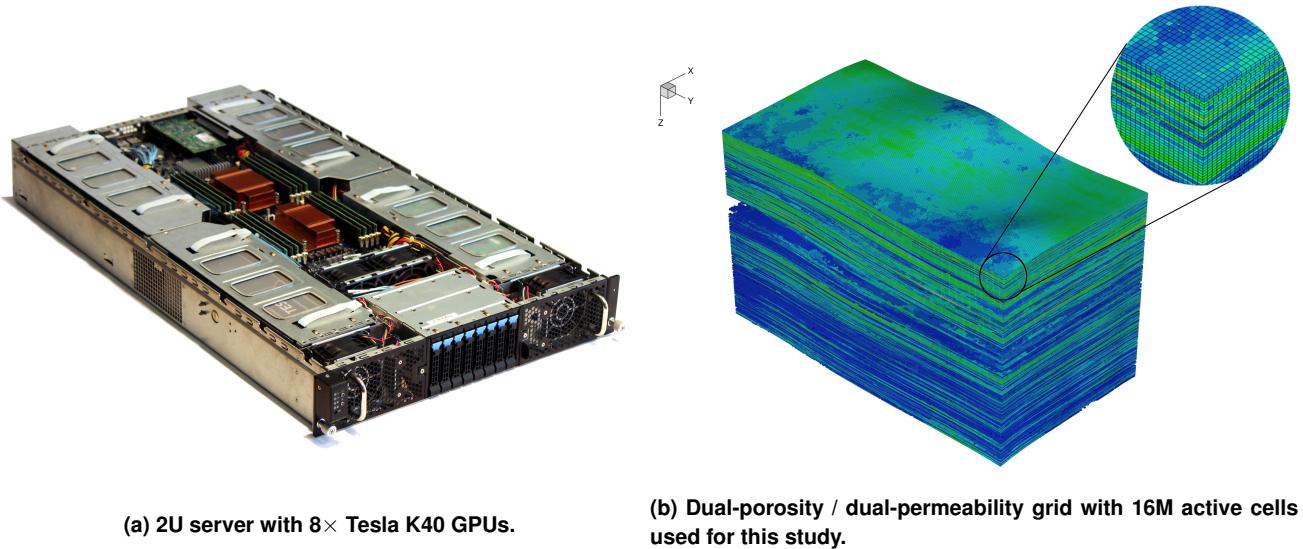


Figure 1: Hardware required to simulate two realizations of a 16 million cell dual-porosity / dual-permeability tight oil model for 20 years of production. Simulation run times are 15 minutes for the dual-porosity and 19 minutes for the dual-permeability cases.

Engineering Workflows for Tight Oil Plays

The study of unconventional tight oil reservoir plays involves three main phases: characterization, history-match, and forecast. The characterization phase involves gathering logs, seismic, and micro-seismic data prior to incorporating them into the geomodel of a reservoir. Matrix characterization includes analysis of well logs calibrated to cores to estimate porosity, fluid saturations, and matrix permeability. Properties are distributed in 3D to honor well tops, geologic understanding, and seismic attributes in order to define internal faulting, possible fracturing, and rock characteristics. Characterization of natural fractures requires calibration to image logs, correlation to seismic attributes, and calibration of these attributes to well performance. The stimulation of natural fractures can lead to areas of locally enhanced permeability called stimulated reservoir volumes (SRVs). The dimensions and permeability of SRVs must be calibrated to hydraulic fracture treatment data, microseismic data, rock characteristics, and early well performance. In the history-matching phase, many simulations of the characterized reservoir geomodel are performed, varying the properties of the geomodel until a good agreement with the known production data is obtained. Finally, in the forecasting phase, oil recovery is predicted using simulations of the history-matched model and often incorporates uncertainty assessment via a large number of simulations.

History-matching and forecasting workflows have different performance requirements. Some aspects of the work are iterative in nature and require fast simulation times for individual runs. Others require high throughput over a large number of independent realizations. Hence, reservoir simulators that satisfy both these requirements play a critical role in accelerating such workflows. The performance provided by a GPU simulator can make such workflows practical for the short time scales demanded by unconventional plays.

Description of the Reservoir Model

The Permian Basin located in West Texas and Southeast New Mexico is the largest onshore oil producing region in the United States. The Permian-age Wolfcamp shale is a well-known source rock present throughout most of the basin and provides opportunities for economic production with hydraulically fractured horizontal wells over a large vertical interval. The Permian Basin index map, Wolfcamp type log and lithofacies are shown in **Fig. 2**. Only the Wolfberry but not the Spraberry interval is included in our model (**Fig. 2b**). The geomodel for the simulation incorporates a number of interpreted petrophysical logs (Fig. 2b) which are used to define

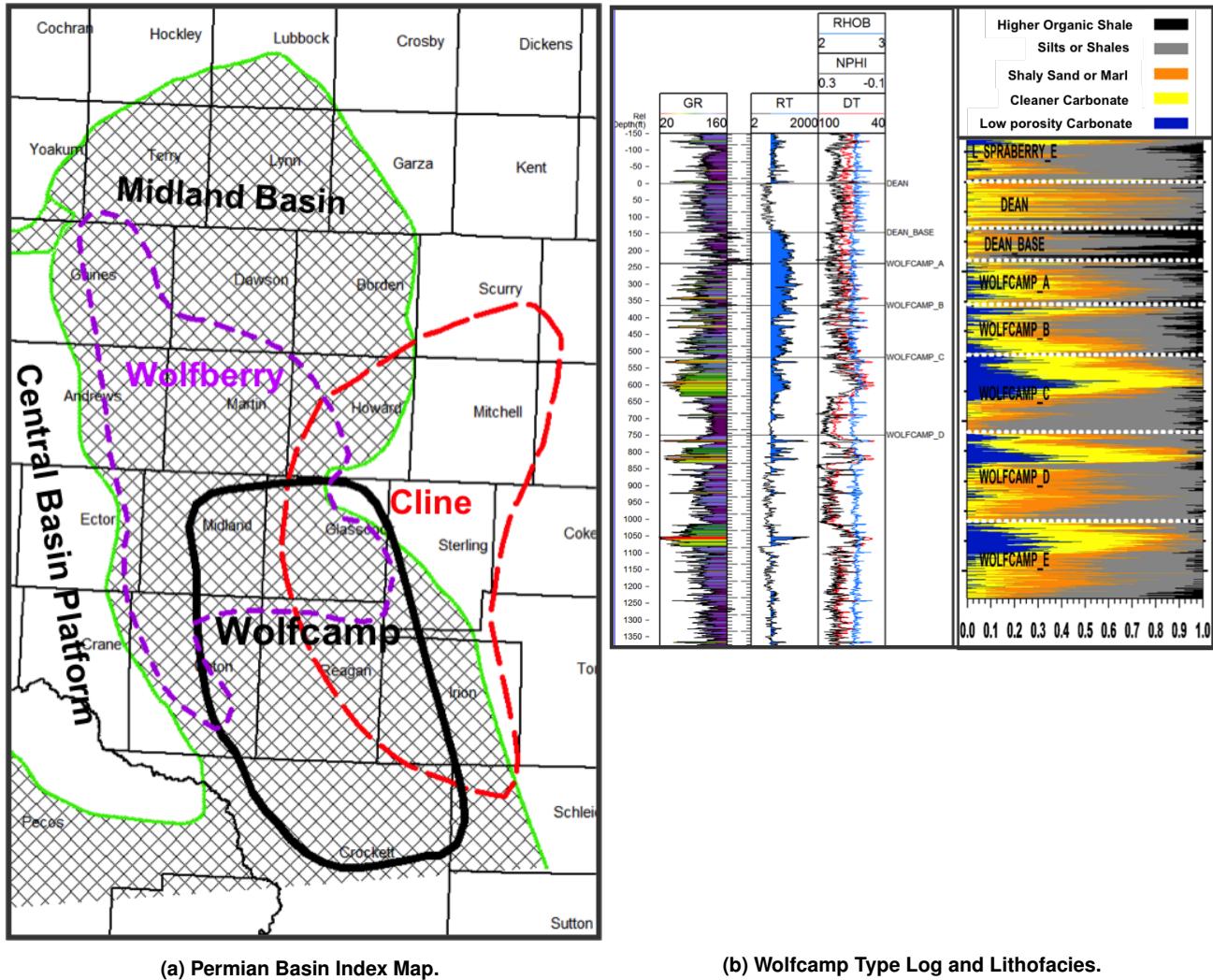


Figure 2: Geologic description of the Permian basin with the Wolfcamp interval.

the vertical distribution of rock matrix, kerogen, porosity, and fluids. These include gamma ray (GR), electrical resistivity (RT), sonic (DT), density (RHOB), and neutron porosity (NPHI) logs. Porosity in the pay interval may be in the range of 6-13%, and kerogen may be on the order of 3-7% of total bulk volume. Kerogen may contain absorbed gas, but in quantities that may be insignificant from a NPV economic point of view and is therefore neglected in the examples shown here. Additional wells with normalized gamma ray logs used as qualitative areal trend guides help control cleaner intervals and shaly trends in the 3D geomodel as a guide for porosity and kerogen distribution away from well control. For the purposes of our simulation, a multi-section geomodel with a fine vertical scale is developed and then upscaled to a 2-mile by 1-mile section with 284 layers, each approximately 4-ft thick. **Fig. 3a** and **Fig. 3** illustrate porosity variations (**Fig. 3a**), fluid saturation variations (**Fig. 3b**), and the vertical zones (**Fig. 3c**) as distributed in the 3D model of the Wolfcamp interval. It can be seen from Fig. 3b that the system contains mobile oil, water and free gas. In the model, propped hydraulic fractures are represented as high permeability dual-porosity regions. This is surrounded by SRVs which approximates the sheared and/or propped natural fractures. The far-field area beyond the SRV is either a single-porosity region (no open natural fractures) or, in the cases shown here, a region with low-permeability unstimulated natural fractures. Dual-media (dual-porosity and dual-permeability) models are used in the numerical simulations to represent these characteristics. The initial average layer properties of the matrix and fracture regions of the Permian Wolfcamp base system are listed in **Table 1**.

Table 1: The initial distribution of layer properties of the base system for the Permian Wolfcamp model.

Parameter	Min. Value	Max. Value	Average Value
Gas Saturation, fraction	0	0.752	0.224
Water Saturation, fraction	0.05	1.00	0.49
Oil Saturation, fraction	0	0.79	0.286
Porosity (for pay only), fraction	0.06	0.13	0.07
Matrix Permeability, nanoDarcy	2.67	125	23.7
Fracture Permeability (unstimulated), nanoDarcy	268	600	434
Fracture Spacing (unstimulated), ft	50	50	50

Because of the large interval and stacked zones with large vertical heterogeneity, the reservoir may not be efficiently depleted without incorporating multiple stacked horizontal wells. **Fig. 3c** shows the 3D model with wells which is the basis for many of the results illustrated in this work. For the model used in this study, a 2 mile by 1 mile area contains 14 horizontal wells each nearly 2 miles in length with over 30 fracture stages per well. For the purposes of discussion, the wells are numbered sequentially starting from the top left of Fig. 3c going to the bottom right. The topmost zone has wells 1-4, followed by the second zone containing wells 5-7 in a staggered manner, then by a third zone containing wells 8-11 and finally by the bottom most zone containing wells 12-14. Nominal well spacing is 1320 ft. in the areal direction, while vertical separation between the centers of the zones is on the order of several hundred feet.

A fine and uniform resolution (40-ft nominal cell width) in the areal direction is used to accurately capture pressure variations and well interactions. We also ran simulation cases with coarser grids. For the coarse grid simulations, the hydraulic fractures are contained in a nominal cell width of 120-ft cells. While we obtained similar results for some of the cases (for example, when the hydraulic fracture and SRV lengths were the same on both grids), the coarser grid severely limited our ability to define alternative fracture lengths as well as areal variability in the geologic characteristics of the matrix and natural fractures. These are important parameters to vary for the sensitivity analysis. For this reason, we retained the finer gridding of the Permian model. Despite the large number of cells, the total computation time for the simulation time was practical.

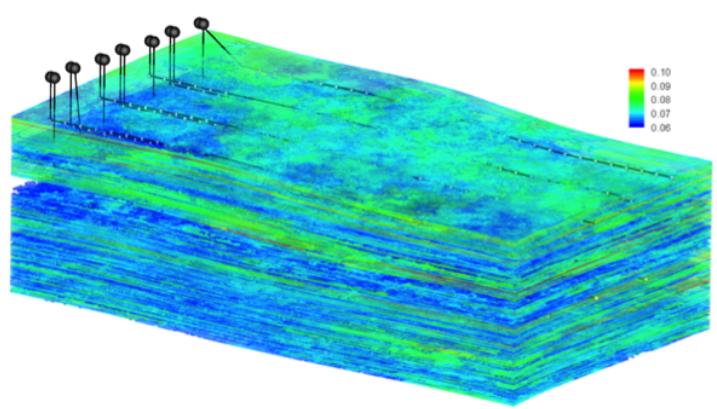
A coarse mesh with local grid refinement (LGR) near the propped hydraulic fractures was also not used because we are more concerned about the accurate estimation of the long term interference rather than the shorter term transient between stages. We have compared results with and without LGRs using commercial simulators. For the types of simulations shown here, the differences in solutions are negligible except at very early times (< 30 days). Besides the fact that propped hydraulic fractures are not ideal bi-wing planes as is usually approximated, the following are the additional reasons LGRs are not required for the present study: (1) the fractures in the stimulated volume allows rapid pressure communication between propped areas. (2) the use of dual-permeability with proper chosen fracture-matrix surface area and directional permeability is equivalent to a single high-permeability discrete grid between low-permeability natural fractures or matrix, and (3) the damage at the face of propped fractures results in behavior similar to a pseudo steady-state approximation over the short time-frames of interest.

Some salient features of the model used for the numerical simulations are listed in **Table 2**.

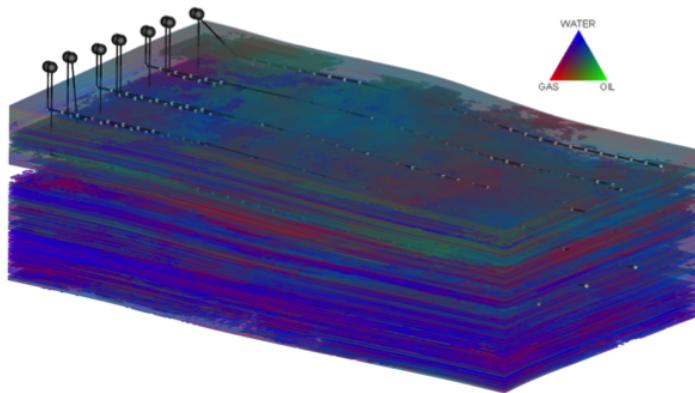
Results And Discussion

For the purposes of validation, we first compare the results from the present simulator with those from a widely used commercial simulator on a model that incorporates the features and properties listed in Tables 1 and 2. The grid parameters for the model are given in **Table 3** along with the absolute run times for these validation cases using the current simulator.

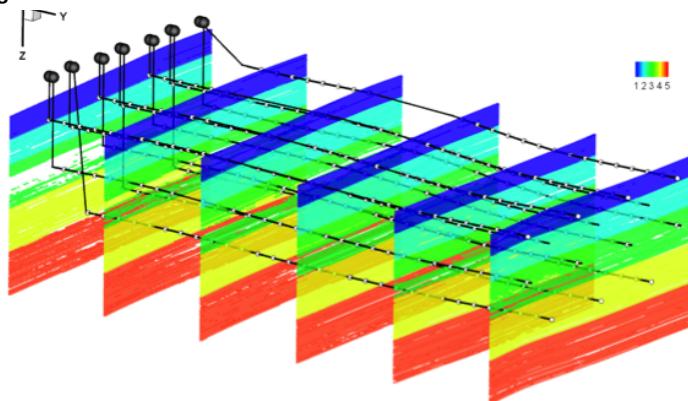
In **Fig. 4**, the results for the cumulative oil productions and field-averaged pressures for the dual-porosity and permeability cases are



(a) Variations of matrix porosity (6% - 13%).



(b) Variations of fluid saturations. The system contains oil, water and gas.



(c) 5 different vertical zones and 14 wells. The topmost zone (dark blue) has wells 1-4 (left to right), the second zone (light blue) has wells 5-7 (left to right), the third zone (green) has wells 8-11 (left to right), and the fourth zone (yellow) contains wells 12-14 (left to right).

Figure 3: Plot of porosity and initial fluid saturation variations in the 2 mile by 1 mile 3D model of the Wolfcamp interval showing a possible multi-stacked well placement. There are 14 stacked and staggered horizontal wells through 5 vertical zones. The vertical scale is exaggerated by a factor of 3 for visualization purposes.

Table 2: Salient features used in modeling the Permian reservoir.

Fracture description	Dual-porosity and dual-permeability
Wells	14 stacked horizontals with multistage fractures
Compaction effects	Dynamic modification of porosity and anisotropic permeability as functions of overburden pressure
Relative permeability	Stone I model for 3-phase relative permeability calculations

Table 3: Simulation parameters for the Permian Wolfcamp model and run times for the base model.

Grid type	Corner-point
Grid dimensions	$132 \times 264 \times 568$
Active cells	16 million
Simulation time	20 years
Time Step	31 days
Run time	15 min. (dual-poro), 19 min. (dual-perm.)

compared with those of the commercial simulator. The insets in **Fig. 4a - Fig. 4d** show the non-normalized differences obtained with both the simulators. The agreement is seen to be excellent and is within 0.01%. Other quantities also show the same level of agreement. The rapid turnaround times as seen from Table 3 for the 16M cell models with complex features indicate the level of acceleration of large model workflows achievable using a single workstation containing 4× Tesla K40 GPUs.

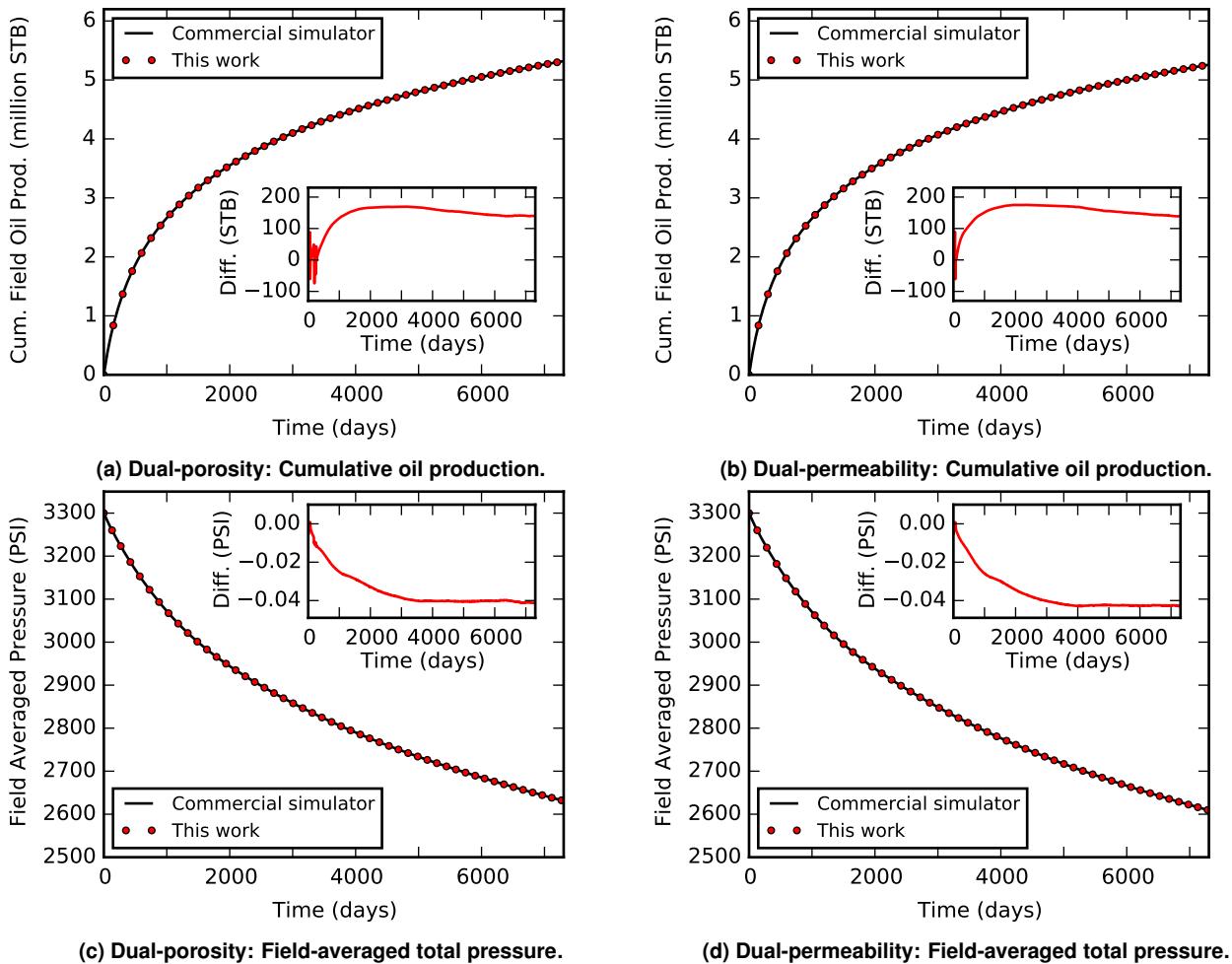


Figure 4: Comparison of dual-porosity and dual-permeability results against a commercial simulator for validation purposes. Insets show the absolute difference in the results between the two simulators. Excellent agreement to within 0.01% is seen for both the dual-media cases.

Next, we demonstrate the importance of numerical simulations to help understand the nonlinear interactions affecting the recovery process. Such an understanding may not be possible using any analytical or empirical methodology due to the inherent approximations used in deriving these models. Consider, for example, the problem of optimizing the number of wells to be drilled for maximal production at minimal cost. A quantitative investigation of the possibilities of long-term well interference is critical for this optimization. To examine these effects, we consider two possible cases for one specific geologic realization of the Permian Wolfcamp basin: (1) all the 14 wells (see, Fig. 3) are open for production, and (2) wells 5-7 in zone 2 and wells 12-14 in zone 4 are shut (or absent).

The results from varying the number of wells are shown in **Fig. 5 – Fig. 6**. For the case in which all 14 wells are open, the pressure contours on the cells in zone 1 containing wells 1-4 are plotted in Fig. 5. **Fig. 5a** shows the contours at an early transient time period. The pressure depletion is localized around each completion and the wells do not interact strongly among each other. However, during a later time period (after 10 years), significant pressure interference can be seen from **Fig. 5b**. In Fig. 6, pressure interference along the vertical direction is highlighted between wells 2 (zone 1) and 9 (zone 3) both during the first year of production (**Fig. 6a**) and after 10 years of production (**Fig. 6b**) which again shows significant interference during later times. The degree of vertical interference will strongly depend on the geomechanical characteristics of the system and hydraulic fracture treatment methods – both of which control height versus areal growth during stimulation. Details on the estimation of hydraulic fracture and SRV sizes are not addressed in this paper. The cumulative oil production plots for both the cases are shown in **Fig. 7** and **Fig. 8**. When all wells are open, the cumulative production decreases compared to the case when some of the wells are either shut or not present. This reduction is more pronounced for wells 8-11 (zone 3) because they strongly interact both with wells 5-7 (zone 2) and wells 12-14 (zone 4), while wells 1-4 (zone 1) strongly interact only with wells 5-7 (zone 2). The insets in Fig. 8 for wells 1 and 2 show the transient behavior for short periods. For short times, particularly for zone 1 wells (1-4), we see that the pressure interferences between vertically stacked layer of wells do not play a major role in the production behavior. But at later times, the plots diverge because of significant interactions among wells. Such behavior cannot be accurately captured by any analytical or empirical models, which are usually developed using a fixed drainage area.

There are inherent uncertainties in defining the characteristics of the geologic and fluid properties. Exploring the sensitivity of recovery forecasts to variations in these parameters require running a large number of independent realizations. In the following, we present results and discussions from multiple realization sensitivity studies of the Permian Wolfcamp interval with 16 million active grid cells. We also give the absolute run times for running a large ensemble of realizations (100) each for the dual-porosity and dual-permeability models. Some of the limitations of the present multiple realization study include: (1) a careful calibration of the interpretations to actual well performance (history matching) has not been completed, (2) the geomodel approximations used here cannot be transferred to all areas of the Wolfcamp play because of the wide variation of properties observed throughout the basin, and (3) the uncertainties in some properties such as relative permeability and compaction have not been included in the sensitivity study.

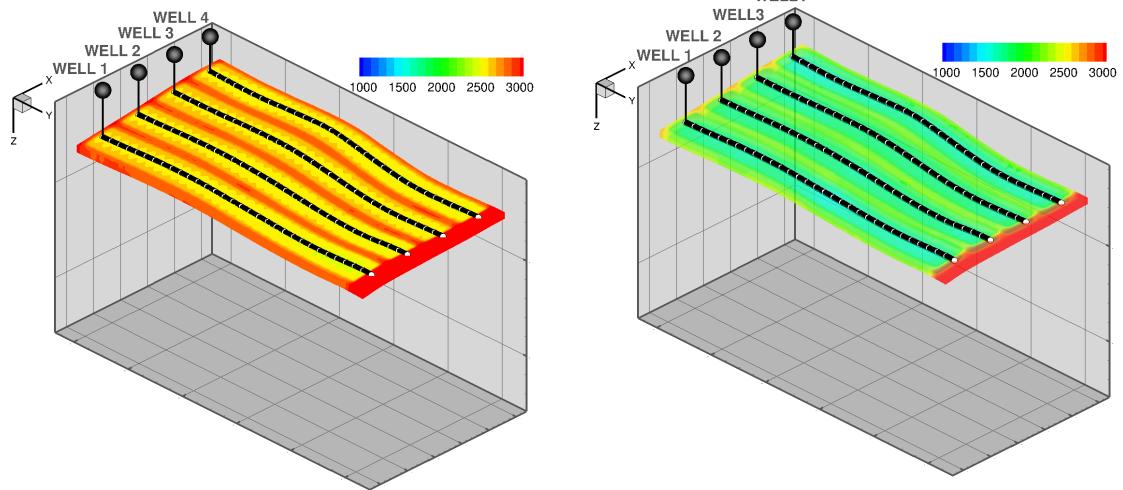
A base case is developed using the properties described in Table 1 and **Table 4**. A schematic of the important geometrical quantities is given in **Fig. 9**. All wells are started at the same time and run with a fixed bottom-hole pressure (BHP) and maximum limiting rate of 500 barrels of oil per day (BOPD) for a forecast period of 20-years. For all the results in the present study, no economic constraints on gas-oil ratio, oil rate, or water-oil ratio have been imposed. Each well has 32-stages (assumed 32 hydraulic fractures). Note that the propped hydraulic fracture is represented by a dual-media idealization of a high permeability fracture in a range of cells corresponding to its length. The hydraulic fractures are contained in a nominal cell width of 40-ft. The effective conductivity is thus given by the fracture effective permeability (k_{frac} , see Table 5) times a cell width of 40-ft ($Frac_w$, see Table 4). However, the effective fracture width (volume) is determined by the assigned fracture porosity of those cells. The multiple realizations are set up as variations on the base case descriptions as listed in Table 5.

Table 4: Dimensions of the fracture and SRV for the base case. Refer to Fig. 9 for the schematic of the fracture and SRV geometries.

Parameter	Value (in ft.)
SRV height (H_{SRV})	90
Fracture height (H_{FRAC})	90
SRV length (L_{SRV})	680
Fracture length (L_{FRAC})	440
SRV width (W_{SRV})	280
Fracture width (W_{FRAC})	40

A linear sensitivity analysis is performed by varying one parameter at a time while keeping all the others fixed at their base values. Using the parameters listed in Table 5, an initial set of 25 dual-porosity realizations was run including the base case. The cumulative run time for all 25 realizations is about 6.5 hours on a single workstation with 4× Tesla K40 GPUs. The variability in the results is evident in **Fig. 10**. This figure shows total oil and gas-oil ratio for Well 1 for a number of realizations. The variation in total oil production for these particular runs is nearly a factor of two for the 20-year forecast (**Fig. 10a**). The total variation considering all 14 wells is even greater. The Gas-oil ratio (**Fig. 10b**) also has significant variability and some unexpected trends.

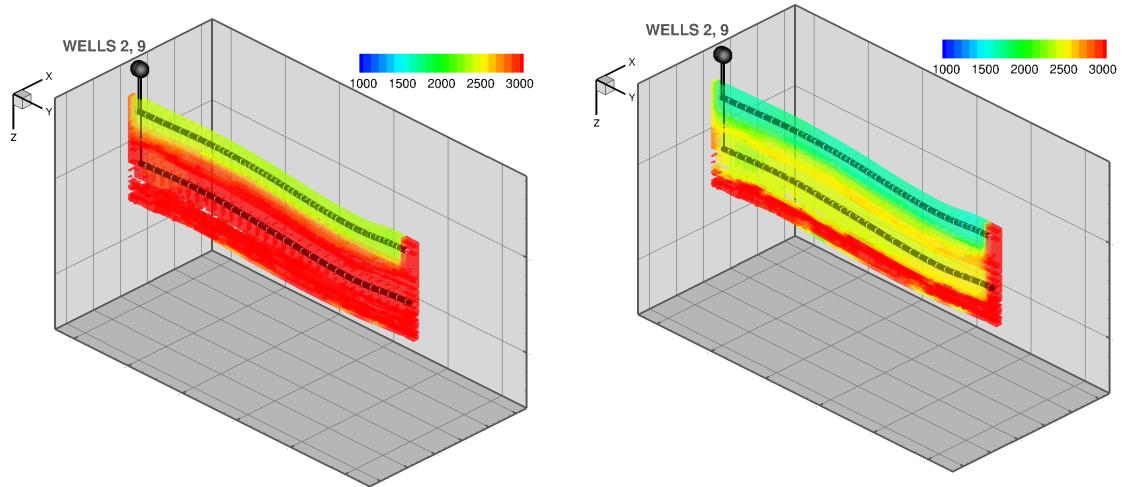
Analysis of decline trends shows significant differences in behavior because of the heterogeneous nature of the system and multi-phase flow effects. Camacho and Raghavan (1989) analyzed Arps' (Arps, 1945) performance-prediction equations under solution-gas



(a) Initial transient pressure distribution.

(b) Pressure distribution after 10-year period of depletion.

Figure 5: The initial transient and long time (10-year depletion period) pressure distributions on areal (I-J) planes containing wells 1–4. Very little interference between wells is seen at the initial period while significant interactions are observed after 10 years of depletion.



(a) Initial transient pressure distribution.

(b) Pressure distribution after 10-year period of depletion.

Figure 6: The initial transient and long time pressure (10-year depletion period) distributions on vertical (J-K) planes containing wells 2 and 9. Very little interference between wells is seen at the initial period while significant interactions are observed after 10 years of depletion.

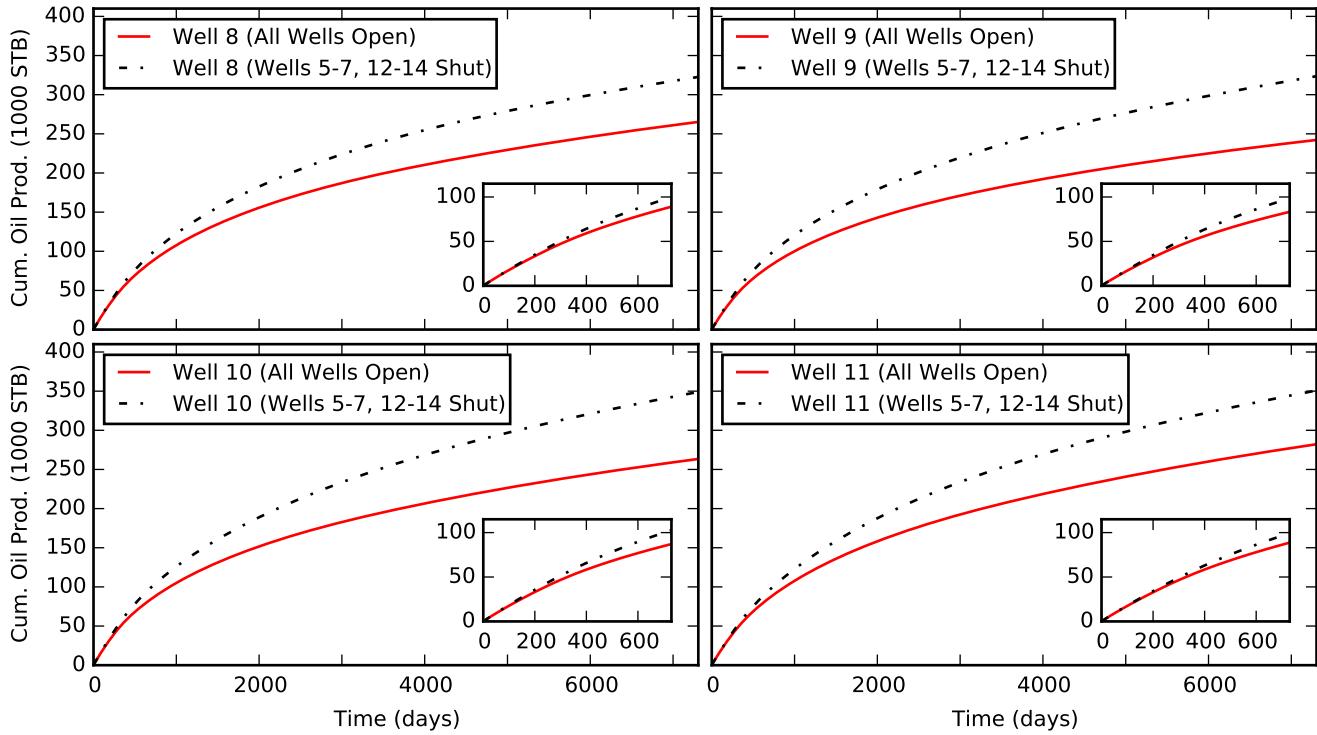


Figure 7: Comparison of cumulative oil production totals for wells 8-11 (zone 3 in Fig. 3) for cases when all wells are open and when wells 5-7 (zone 2), 12-14 (zone 4) are shut. Interference between zones 2 and 4 leads to significantly lower cumulative oil production. Insets show that this interference is apparent even in a shorter two-year window.

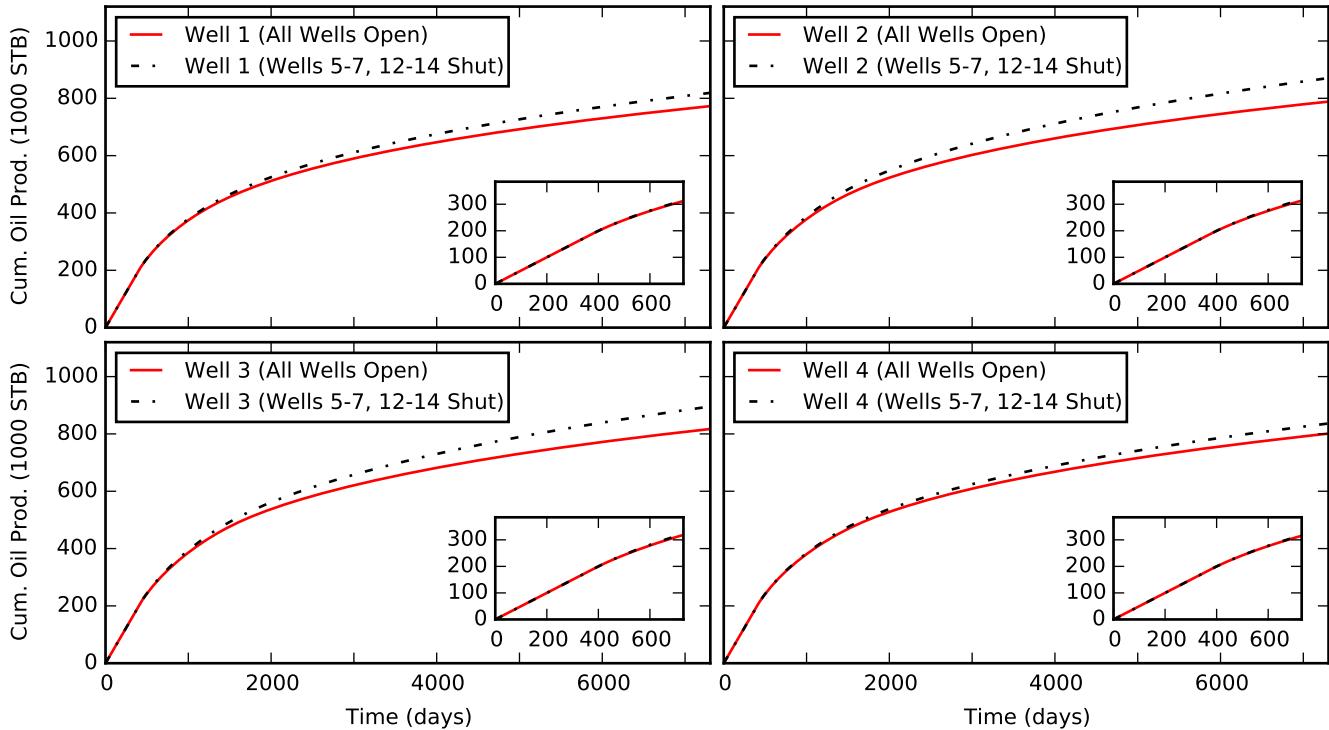


Figure 8: Comparison of cumulative oil production totals for wells 1-4 (zone 1 in Fig. 3) for cases when all wells are open and when wells 5-7 (zone 2), 12-14 (zone 4) are shut. Interference between zone 2 leads to lower cumulative oil production. Insets show that this interference for zone 1 wells is not apparent in a shorter two-year window,

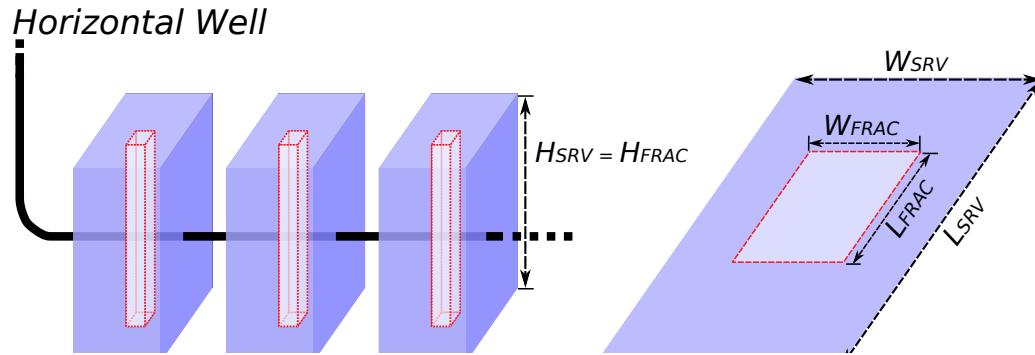
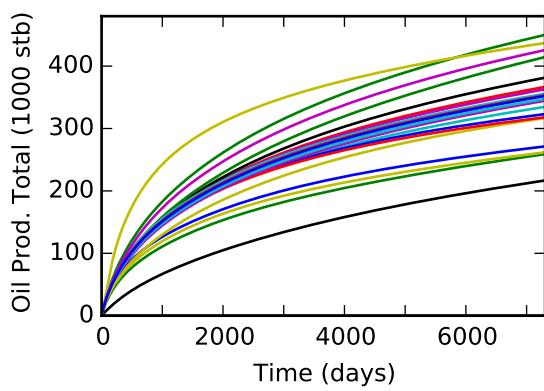
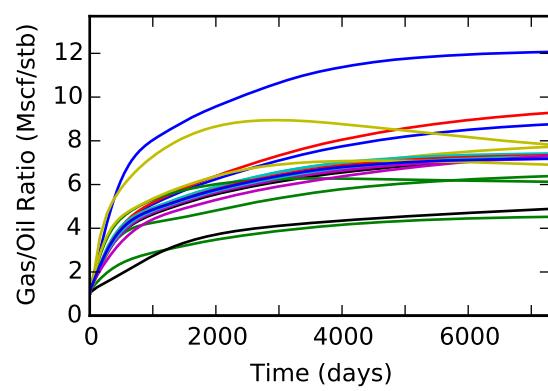


Figure 9: Schematic of hydraulic fractures (*FRAC*) and stimulated reservoir volumes (*SRV*). Variables *L*, *W*, and *H* denote length, width, and height, respectively. The dimensions are not to scale.



(a) Oil production variability for Well 1.



(b) Solution gas-oil-ratio variability for Well 1.

Figure 10: Variation in total oil production and gas-oil ratio for Well 1.

Table 5: Baseline properties for the dual-porosity reservoir model.

Parameter	Min.	Base	Max.
Initial gas saturation (multiplier)	0.8	0.8	1
Average matrix porosity multiplier	0.9	1	1.1
Average matrix permeability multiplier	0.2	1	2.0
Matrix-fracture transmissibility multiplier	0.1	1	10.0
Natural fracture permeability multiplier (unstimulated areas)	0.1	1	10.0
SRV permeability (in mD)	0.05	0.1	0.2
SRV height (H_{SRV} in ft.)	56	90	122
SRV areal extent (L_{SRV} in ft.)	440	680	920
Fracture height (H_{FRAC} in ft.)	56	90	122
Fracture areal extent (L_{FRAC} in ft.)	280	440	600
Fracture Permeability (k_{frac}) (in mD)	0.25	1.25	6.25
Well spacing (refer Fig. 3c)	Wells 5-7, 12-14 Shut	All 14 wells Open	Wells 1-4, 8-11 Shut

drive in conventional reservoirs showing the importance of parameters like relative permeability, PVT data, and pore volume on the prediction of long-term performance. Consider the 3-parameter Arps' equation given by **Eqn. 1**. Here, q is the instantaneous rate, q_i is the initial rate, b is an exponential constant and D_i is the nominal decline rate.

$$q = q_i(1 - bD_i t)^{-1/b}. \quad (1)$$

The hyperbolic decline equation with constant b is obtained for a linear variation of fluid mobility to total system compressibility in Eqn. 1. For multiphase reservoirs, the mobility/compressibility ratio function becomes nonlinear with time and reservoir characteristics vary over the drainage volume, which lead to a non-constant b . Camacho and Raghavan (1989) noted that if transient data is included in Arps' decline curves, b will be a function of time with values greater than unity in most of the cases. Kurtoglu et al. (2011) discussed similar considerations with regard to long term performance of oil wells in Elm Coulee Field (Bakken) and illustrated how the long term performance of one well could be fit by b factors ranging from 2.0 for the first six months, then 1.9 through 12 months and 1.1 after 24 months.

Fig. 11 shows hyperbolic decline equations (Eqn. 1) fit to base-case and a high-side simulation (lower initial gas saturation) case for Well 1. Early time declines are best fit with higher b and higher nominal decline (D_i) as expected for transient behavior (dashed lines). These fits to early time data over-predict long-term recovery; however, the long term decline still shows a hyperbolic behavior because of multiphase flow. The variation from the high-side to low side case is a complex function of the reservoir properties, stimulation characteristics, fluid behaviors and possible well interference which are best estimated via simulation using an integrated characterization effort and uncertainty assessment. Other decline curves fit to simulation results are illustrated by Gilman and Fadaei (2013); however for this paper we retain use of the Arps equation because of its wide use and familiarity.

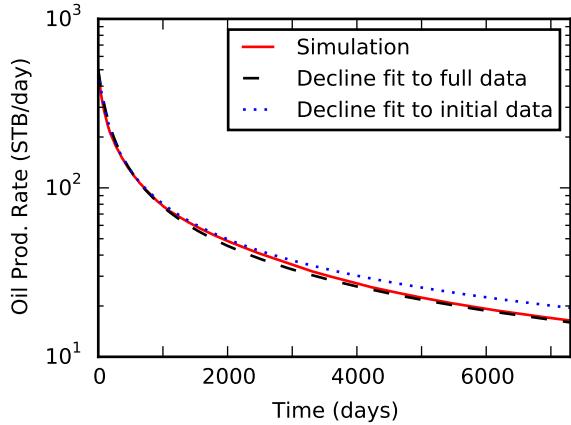
Simulation Times for Multiple Realizations

Multiple realization were run with both dual-porosity and dual-permeability formulations. These realizations were generated using a quasi-random Sobol sequence (Sobol, 1967) using the ranges specified in Table 5. For demonstration purposes, a total of 100 realizations were generated each for the dual-porosity and dual-permeability cases. The total wall clock times for running these realizations on a single workstation with 4× Tesla K40 GPUs are shown in (**Table 6**). For the cases simulated, both the dual-porosity and dual-

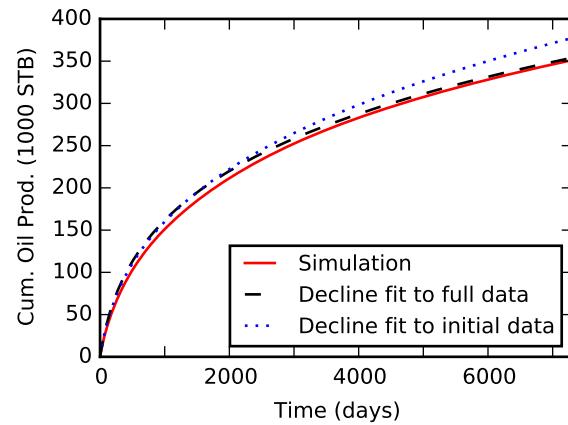
Table 6: Cumulative run times and average run times per simulation for multiple realizations of the dual-media Permian Wolfcamp model with 16 million active cells and parameters derived from Table 5. The realizations were run on a single workstation with 4× Tesla K40 GPUs.

Dual-medium model	Number of realizations	Cumulative run time	Average run time / simulation
Dual-porosity	100	26.25 hrs.	15.9 mins.
Dual-permeability	100	34 hrs.	20.4 mins.

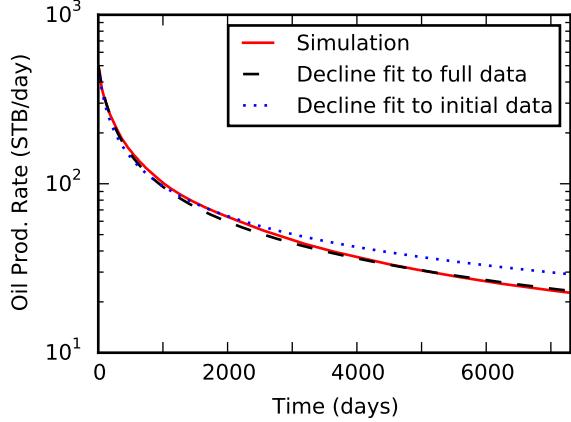
permeability gave similar results. This is because of the high contrast between fracture and matrix permeability. However, the total computational times to solve both the dual-porosity and dual-permeability formulations are not very different. From the simulation times, it is clear that a large number of full-physics high fidelity realizations can be achieved in a reasonable amount of time using a GPU-accelerated simulator. Using a single rack consisting of 160 Tesla K40 GPUs in 20 nodes, it is possible to *simultaneously* simulate 40 realizations in 16 minutes for dual-porosity and in 20 minutes for dual-permeability models. Based on preliminary tests using the latest Tesla K80 GPUs, the number of simultaneous realizations can be doubled with very similar run times. It should be noted that our results for timings are over an order-of-magnitude faster for the dual-porosity cases and nearly 20-times faster for the dual-permeability



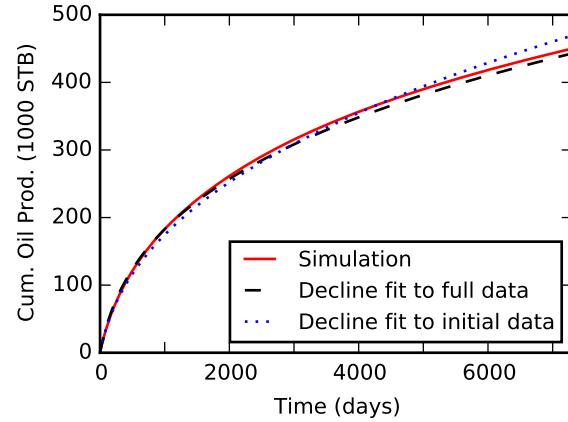
(a) Base simulation case (blue line) oil production rate for Well 1. Early time fit (red line) is matched with a $b = 1.35$ and $D_i = 94.6\%$. The best fit for 20-years (green line) is $b = 1.2$ and $D_i = 92.2\%$.



(b) Base simulation case (blue line) oil production total for Well 1. Early time fit (red line) is matched with a $b = 1.35$ and $D_i = 94.6\%$. The best fit for 20-years (green line) is $b = 1.2$ and $D_i = 92.2\%$.



(c) Low-side (lower initial gas saturation) simulation case (blue line) oil production rate for Well 1. Early time fit (red line) is matched with a $b = 1.6$ and $D_i = 94.6\%$. The best fit for 20-years (green line) is $b = 1.33$ and $D_i = 88.8\%$.



(d) Low-side (lower initial gas saturation) simulation case (blue line) oil production total for Well 1. Early time fit (red line) is matched with a $b = 1.6$ and $D_i = 94.6\%$. The best fit for 20-years (green line) is $b = 1.33$ and $D_i = 88.8\%$.

Figure 11: Hyperbolic decline equation (Eqn. 1) fit to the base- and low-side simulation (lower initial gas saturation) cases. Note that for both the cases, early time match over predicts recovery, but the long term decline still shows hyperbolic behavior because of multiphase flow.

cases when compared to a commercial simulator running on 12 CPU cores with recent processors. As the properties and gridding become more heterogeneous, even more speedup (two orders of magnitude) has been noted (Esler et al., 2014).

Summary

Numerical models are required to assess complex interactions of pressure interference, multi-phase flow and detailed reservoir heterogeneities. Geologic models that honor features such as stacked pay intervals, and completion complexities such as long well length, numerous hydraulic stages, and closely spaced wells require grid sizes that run into many millions of cells. Coupled to this is the uncertainty factor in defining the characteristics of the geologic and fluid properties in order to forecast recovery which requires running multiple realizations for sensitivity analysis and history matching these large models. Using representative results, we highlight the significant impact that the realistic geologic and fluid models can have on the decline curves which cannot be fully addressed by simple empirical and analytical equations.

One of the main advantages of a fully optimized GPU formulation described here is accelerating complex workflows by performing full-physics simulations of the black-oil system involving multiple realizations of models with many millions of cells. This is mainly possible by using state-of-the-art algorithms optimized for use with multiGPUs, addressing Amdahl's law, and through careful and efficient memory management. Current CPU formulations generally cannot handle such heavy workloads in a reasonable amount of time for dual-porosity and especially dual-permeability models. While a cluster-based CPU solution can address some of these needs there is often large overhead that increases the computational time.

We demonstrated efficient GPU capability through numerical simulations of a real-world unconventional play, namely, Permian basin characterization of the Wolfcamp interval developed with stacked horizontal wells and multi-stage fractures. Both dual-porosity and dual-permeability modeling is used in the present study with up to 16 million active cells. We provide timing results from performing hundreds of simulations (choosing a representative set of a few hundred samples from the sensitivity parameter space) to illustrate the speedup possible from using the fine-grained GPU simulator. To the best of our knowledge, for the first time, we have convincingly demonstrated that large numbers of realizations on models involving tens of millions of cells can be performed in realistic timeframes and using only modest hardware requirements.

Acknowledgments

We would like to thank the iReservoir staff for the detailed data analysis, characterization and geomodeling required to develop the reservoir description and upscaling used in the model. We especially wish to acknowledge the efforts of Mike Uland and Omar Angola. We would also like to thank NVIDIA for lending GPUs and allowing access to their cluster hardware.

Nomenclature

b =	hyperbolic decline parameter in Arps' equation, dimensionless
D_i =	initial decline term in Arps equation, 1/day
q =	flow rate, STB/D
q_i =	initial flow rate in Arps' equation, STB/D
t =	time, days
k =	permeability, milliDarcy
W =	width, ft
H =	height, ft
L =	length, ft
SRV =	stimulated reservoir volume, ft^3

Subscripts

$FRAC$ =	fracture, dimensionless
SRV =	stimulated reservoir volume, dimensionless

References

- J. Appleyard, J. Appleyard, M. Wakefield, and A. Desitter. Accelerating Reservoir Simulators using GPU Technology. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2011. ISBN 978-1-55563-324-0. doi: 10.2118/141402-MS.
- J. J. Arps. Analysis of Decline Curves. *Transactions of the AIME*, 160(01):228–247, 1945.
- M. Bayat and J. E. Killough. An Experimental Study of GPU Acceleration for Reservoir Simulation. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2013. ISBN 978-1-61399-233-3. doi: 10.2118/163628-MS.

- M. Brown, E. Ozkan, R. Raghavan, and H. Kazemi. Practical solutions for pressure-transient responses of fractured horizontal wells in unconventional shale reservoirs. volume 14, pages 663–676. Society of Petroleum Engineers, 2011.
- R. Camacho and R. Raghavan. Boundary-Dominated Flow in Solution Gas-Drive Reservoirs. In *Low Permeability Reservoirs Symposium*, number SPE-19009-MS. Society of Petroleum Engineers, 1989.
- Z. J. Chen, H. Liu, and S. Yu. Development of Algebraic Multigrid Solvers Using GPUs. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2013. ISBN 978-1-61399-233-3. doi: 10.2118/163661-MS.
- K. Esler, K. Mukundakrishnan, V. Natoli, J. Shumway, Y. Zhang, and J. Gilman. Realizing the Potential of GPUs for Reservoir Simulation. In *ECMOR XIV-14th European conference on the mathematics of oil recovery*, 2014.
- K. P. Esler, V. Natoli, and A. Samardzic. GAMPACK (GPU Accelerated Algebraic Multigrid Package). In *ECMOR XIII-13th European Conference on the Mathematics of Oil Recovery*, 2012.
- D. Foltinek, D. Eaton, J. Mahovsky, P. Moghaddam, and R. McGarry. Industrial-scale reverse time migration on GPU hardware. In *SEG Houston International Exposition*, 2009.
- J. R. Gilman and S. Fadaei. Non-Linearity in Decline Behavior of Unconventional Wells. Unconventional Resources Technology Conference (URTEC), 2013.
- B. Kurtoglu, S. A. Cox, and H. Kazemi. Evaluation of Long-Term Performance of Oil Wells in Elm Coulee Field. In *Canadian Unconventional Resources Conference*. Society of Petroleum Engineers, 2011.
- M. Lin, S. Chen, Z. Chen, and J. Xu. Fractured Reservoir Modeling: Effects of Hydraulic Fracture Geometries in Tight Oil Reservoirs. In *SPE/EAGE European Unconventional Resources Conference and Exhibition*, number SPE-167761-MS. Society of Petroleum Engineers, 2014.
- B. A. Ogunyomi, T. W. Patzek, L. W. Lake, and C. S. Kabir. History Matching and Rate Forecasting in Unconventional Oil Reservoirs Using an Approximate Analytical Solution to the Double Porosity Model. In *SPE Eastern Regional Meeting*, number SPE-171031-MS. Society of Petroleum Engineers, 2014.
- V. Okouma, D. Symmons, N. Hosseinpour-Zonoozi, D. Ilk, and T. A. Blasingame. Practical Considerations for Decline Curve Analysis in Unconventional Reservoirs—Application of Recently Developed Rate-Time Relations. In *Paper SPE 162910 presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Calgary, Alberta, Canada*, pages 24–25, 2012.
- I. M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.
- E. Stalgorova and L. Mattar. Analytical Model for History Matching and Forecasting Production in Multifrac Composite Systems. In *SPE Canadian Unconventional Resources Conference*, number SPE-162516-MS. Society of Petroleum Engineers, 2012a.
- E. Stalgorova and L. Mattar. Practical analytical model to simulate production of horizontal wells with branch fractures. In *SPE Canadian Unconventional Resources Conference*, number SPE-162515-MS. Society of Petroleum Engineers, 2012b.
- Hamdi Tchelepi and Yifan Zhou. Multi-GPU parallelization of nested factorization for solving large linear systems. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2013.
- S. Yu, H. Liu, Z. J. Chen, B. Hsieh, and L. Shao. GPU-based Parallel Reservoir Simulation for Large-scale Simulation Problems. In *SPE Europe/EAGE Annual Conference*. Society of Petroleum Engineers, 2012. ISBN 978-1-61399-204-3. doi: 10.2118/152271-MS.