MODEL DOCUMENTATION

RIO Model

United States

Date	October 26, 2026
Package	EnergyPATHWAYS and RIO Documentation, October 26, 2026

Contents

1 Model framework

- 1.1 Scope
- 1.2 Demand modes
- 1.3 Inputs read from EP
- 1.4 Geographic scope, horizon, and units
- 1.5 Notation

2 Temporal resolution, solution algorithm, and objective

- 2.1 Temporal aggregation
- 2.2 Solution algorithm
- 2.3 Objective function

3 Capacity investment and retirement

- 3.1 Capacity accounting
- 3.2 Capacity pathways
- 3.3 Storage duration
- 3.4 Resource potential and build-rate limits
- 3.5 Portfolio and retrofit constraints

4 Operations

- 4.1 Electricity balance
- 4.2 Thermal generation
- 4.3 Fixed-profile generation
- 4.4 Hydro generation
- 4.5 Electric storage
- 4.6 Planning reserves
- 4.7 Operating reserves
- 4.8 Transmission
- 4.9 Network peak and distribution cost attribution

5 Sector coupling

- 5.1 Carrier balances
- 5.2 Carrier storage and transport
- 5.3 Conversion technologies
- 5.4 Commodity supply and industrial materials
- 5.5 Emissions accounting

6 Hydrocarbon supply representation

- 6.1 Resource discretization
- 6.2 Breakeven supply cost
- 6.3 Extraction and processing technologies
- 6.4 Drilling capacity constraints
- 6.5 Intra-bin cost escalation
- 6.6 Drilling-pace price response

6.7 Project ramp constraint
6.8 Midstream, trade, and upstream emissions

7 Sectoral carbon-intensity standards

7.1 Product group definition
7.2 Constraint formulation
7.3 Emissions attribution
7.4 Product supply accounting
7.5 Deemed emission intensities and trading

8 Demand representation and demand-side optimization

8.1 Exogenous demand
8.2 Load shifting
8.3 Decision scope in mode (b)
8.4 Adoption bounds
8.5 Sales and service demand
8.6 Payback-acceptance allocation
8.7 Technology linkage and ratio constraints
8.8 Efficiency and fuel-switching measures
8.9 Supply-side coupling and cost accounting
8.10 Mode-specific variables and constraints

References

This document is the documentation of the RIO model as applied to the United States energy system. RIO is a linear program that optimizes investment in and operation of energy supply, together with the demand-side technology choices left open by the EnergyPATHWAYS demand model. The document describes what the model represents and how: the decision variables, constraints, and objective, the two modes in which demand enters the program, and the features available to a run. A particular run selects data, parameter values, geography, model years, temporal resolution, and the features that are switched on. The data and settings for particular runs are documented in the *RIO US Database* document. Williams et al. (2021) give an earlier description of the framework.

1 Model framework

1.1 Scope

RIO is a linear program of energy supply. It co-determines, in a single optimization, investment and operation across electricity, energy carriers, commodities, and industrial materials. When demand optimization is switched on (Section 1.2), the same program also determines demand-side technology adoption within bounds supplied by EnergyPATHWAYS (Section 8). Formulating supply in a single program allocates resources with competing end uses on a common marginal basis: biomass between liquid fuels and generation, hydrogen between industrial heat and synthesis, CO_2 sequestration capacity between capture sources, and low-carbon electricity between direct end use and fuel production.

EnergyPATHWAYS (EP) is a separate stock-rollover accounting model of energy demand. It projects energy service demand, equipment stocks, final energy demand by carrier, hourly load profiles, and demand-side cost for each demand subsector, and RIO reads its outputs. EP is documented in the *EnergyPATHWAYS Model* document. The models are run in series: EP passes demand quantities, profiles, technology characterization, and costs to RIO, and no prices are passed back to EP.

The level of energy service demand is exogenous to RIO. RIO represents no income effect, no price-induced change in service demand, and no macroeconomic feedback. Where RIO changes demand, it does so by substituting technologies or measures at constant service demand.

1.2 Demand modes

Demand enters RIO in one of two modes, selected in the run configuration by a demand optimization switch for demand technologies and a second switch for efficiency and fuel-switching measures.

(a) Demand optimization off. RIO takes the demand supplied by EP as fixed. Final energy demand enters the balances as exogenous quantities: an hourly electricity load by zone, sector, and distribution feeder; annual demand for every other carrier by zone, sector, and year; and hourly profiles for those other carriers that are balanced at sub-annual resolution. Every subsector is treated as exogenous, and no demand technology data are read. Efficiency and fuel-switching measures are held at their adoption in the EP case, which is already reflected in the exogenous demand. RIO optimizes supply only. The demand terms of Section 8.1 are the whole of demand in the balances. Load shifting (Section 8.2) is the only demand-side flexibility in this mode, and it applies where flexible-load input data are supplied.

(b) Demand optimization on. RIO chooses sales of demand-side technologies for the subsectors designated for optimization, within an adoption envelope supplied by EP, and chooses the adoption of efficiency and fuel-switching measures. Exogenous demand from EP remains in the balances for all subsectors. For optimized subsectors it is the demand of the EP baseline sales, and RIO adds an endogenous term equal to the change in demand caused by departures from that baseline (Sections 8.3 to 8.9). Subsectors not designated for optimization are handled as in mode (a).

The demand technology switch takes effect only when demand technology cost data are present in the run inputs. The measure switch is used together with the technology switch, because the measure cost accounting and the non-decreasing condition of Section 8.8 belong to the demand-optimization constraint set. The variables and constraints that exist only in mode (b) are listed in Section 8.10.

1.3 Inputs read from EP

RIO reads five categories of input from EP.

Category	Contents	Mode
Exogenous demand	Annual final energy demand by subsector, sector, carrier, and zone; hourly electricity load by zone, sector, and distribution feeder for each weather year; hourly profiles for other carriers requiring sub-annual resolution; distribution feeder allocation	(a) and (b)
Flexible load	Flexible fraction of subsector load and load-shifting cost	Where supplied
Technology characterization	Baseline, minimum, and maximum sales; service demand per unit sales; electric and fuel efficiency; hourly profiles; balance resolution; capital cost, fixed operating cost, and incentives; lifetimes; technology linkages and ratio constraints; payback-acceptance shares, by technology, vintage, and zone	(b) only
Demand-side measures	Efficiency measure savings and cost; fuel-switching measure savings, additions, and cost	Measure switch on
Cost accounting	Levelized demand-side cost	Reporting only

In mode (b), exogenous demand for an optimized subsector is the demand of its EP baseline sales, and the technology characterization defines how sales can depart from that baseline. Both models resolve identical zones and weather years.

1.4 Geographic scope, horizon, and units

The model covers the contiguous United States and can represent Alaska and Hawaii as separate zones. RIO resolves a set of model zones, together with one non-physical zone carrying international trade flows. Zones are assembled from counties, and zone systems at state-level resolution are supported. Input data are entered in their native geographies and mapped to the model zones by allocation keys. Several constraint families are enforced over their own geographies, coarser or finer than the model zones, mapped by allocation ratio: drilling capacity over drilling regions and a national aggregate (Section 6.4), and sectoral carbon-intensity standards over cap zones defined in the input data (Section 7.2).

RIO is solved for a set of optimization years, each carrying a gap weight g_y equal to the number of calendar years it represents.

The model works on a single unit and currency basis set in the run configuration. Costs are discounted at a real discount rate from a base year. Demand-side costs read from EP are converted to the RIO currency basis before they are combined with supply-side costs.

1.5 Notation

Symbol	Set	Description
y	$\mathcal{Y}$	Optimization years
z, q	$\mathcal{Z}$	Model zones
s	$\mathcal{S}$	Representative days
t	$\mathcal{T}$	Hours within a representative day
d	$\mathcal{D}$	Period blocks in the annual chronology
ω	$\mathcal{W}$	Capacity-reservation windows
i, n		Supply technology, operating mode
k		Electric storage technology
x		Transmission connection
c, b		Commodity, supply-curve bin
β	$\mathcal{B}$	Energy carriers and material products
g	$\mathcal{G}$	Greenhouse gases
w, j		Demand sector, distribution feeder
g, τ		Demand group, demand technology
m		Efficiency or fuel-switching measure
p	$\mathcal{P}$	Payback-acceptance bins
γ, ζ		Product group, cap zone
g_y		Gap weight of optimization year y
$w_{y,s}$		Calendar days represented by representative day s
δ_y		Discount factor, $(1-r)^{\,y-y_0}$ for discount rate r and base year y_0
W		Number of weather years

2 Temporal resolution, solution algorithm, and objective

2.1 Temporal aggregation

The year is represented by a set of representative days, weighted to 365 calendar days. The number of representative days is a run setting. Each calendar day is characterized by features computed predominantly from net load. Net load is computed for several renewable portfolios, each sized so that renewable generation equals annual load energy: one portfolio per renewable resource, with 80% of the energy from that resource and the remainder split among the others, plus one portfolio with equal shares. Days are clustered in that feature space by agglomerative clustering with complete linkage. Each cluster is represented by the member day that minimizes the summed squared difference in daily net-load energy across the cluster, so each representative day is an observed historical day. Each representative day s carries weight $w_{y,s}$ equal to the number of calendar days it represents, and hourly operating costs are scaled by $w_{y,s}$. Clustering on net load conditions the representative set on the ramp and over-generation characteristics that determine integration cost. Representative days whose daily net-load energy or peak exceeds a configurable percentile are flagged as reliability-

critical, and the planning reserve constraint of Section 4.6 is enforced on those days. A percentile of zero flags every representative day. Load and renewable generation profiles are taken from W weather years, the same weather years used for the EP load profiles.

Representative days are ordered within the year by a period layer. Each calendar day is mapped to its representative day, and consecutive calendar days sharing a representative day are aggregated into a weighted period block d. Period blocks in calendar order constitute an approximate annual chronology. A capacity-reservation window layer ω aggregates period blocks seasonally; the window length is a run setting.

Operational variables are resolved at one of three timescales, specified per technology and per carrier: hourly (H), representative-day (S), or annual (Y). Fast resources are resolved hourly; slow resources, including annual fuel production, are balanced annually. Terms at finer resolution are aggregated into coarser balances by the representative-day weights.

2.2 Solution algorithm

RIO is solved as a sequence of receding-horizon passes, one per optimization year. Each pass optimizes the years within its foresight window and fixes investment from preceding passes. Committed capacity is transferred forward between passes, and investment does not anticipate conditions beyond the current window. The window length is configurable. A window that contains only its own optimization year makes each pass myopic. The program is a linear program, and the solver and its settings are run options.

Policy constraints carry penalized slack variables in place of hard limits, so an infeasible policy specification yields a solution with identifiable slack instead of an infeasible program. Unserved energy is admitted at a value-of-lost-load penalty.

In mode (b), the demand-side constraints are first solved on their own with an objective equal to the summed slack on the adoption envelope (Section 8.4). This solve checks that the envelope is feasible. Its variable values are discarded before the full program is solved.

2.3 Objective function

The objective minimizes discounted total system cost, assembled from named components each defined as an expression filterable by the years active in a solution pass and as a constraint on a named cost variable.

Group	Components
Supply-side capital	New build, life extension, repower, and retrofit capital for generation and storage; retirement capital for generation; interconnection cost; investment and production tax credits
Supply-side operating	Fixed and variable operating cost; startup and shutdown cost; curtailment cost
Carriers and commodities	Commodity resource cost; intra-bin depletion cost; drilling-pace rent; commodity and carrier delivery cost; conversion delivery cost
Networks	Transmission capital and fixed operating cost; hurdle-rate cost; system and feeder peak cost
Demand side	Electricity delivery cost; load-shifting cost; in mode (b), demand technology capital cost, fixed operating cost, and incentives, and efficiency and fuel-switching measure cost
Policy	Emissions price cost; clean-energy standard prices, incentives, offsets, and trading friction

Group	Components
Penalties	Unserved energy; unmet capacity reserves; over-generation; hydro budget violation; early and forced retirement; growth; annual build-limit exceedance; storage state of charge; and balance, allocation, cap, envelope, and adoption-condition slacks

Each component is multiplied by the discount factor δ_y of its year. Capital costs enter as annualized charges. Operating costs are annualized by the representative-day weights of Section 2.1. Drilling-pace rent (Section 6.6) is booked separately from commodity resource cost as a transfer to resource owners, and total cost is reportable with or without it. Technologies with annual build limits also carry a small charge on annual build, $0.01/(y - y_0 + 1)$ per unit, which breaks ties between otherwise equal-cost build years.

In mode (a) the objective contains no demand technology or measure cost, because the demand side is fixed. The levelized demand-side cost read from EP is carried for reporting in both modes and does not enter the objective.

3 Capacity investment and retirement

3.1 Capacity accounting

Online capacity is an accounting expression, not a free variable. For technology i in year y, capacity sums five contribution streams, each windowed by the lifetime of its pathway:

$$\begin{aligned}\mathrm{Cap}_{y,i} \;=\; &\sum_{v \in V^{\mathrm{new}}_{y,i}} \Big(\mathrm{New}_{v,i} - \sum_{q \le y} \mathrm{RetN}_{q,v,i}\Big) + \sum_{v \in V^{\mathrm{rtf}}_{y,i}} \Big(\mathrm{Retrofit}_{v,i} - \sum_{q \le y} \mathrm{RetR}_{q,v,i}\Big) \\ &+ \sum_{v \in V^{\mathrm{ext}}_{y,i}} \mathrm{Extend}_{v,i} + \sum_{v \in V^{\mathrm{rep}}_{y,i}} \mathrm{Repower}_{v,i} + \sum_{e \in X_{y,i}} \Big(\mathrm{Exist}_{e} - \sum_{q \le y} \mathrm{RetE}_{e,q}\Big) + \sigma^{+}_{y,i} + \sigma^{-}_{y,i}\end{aligned} \tag{1}$$

$V^{\mathrm{new}}_{y,i}$ and the other vintage sets contain the build years v with $y - L < v \le y$ for the lifetime L of the pathway, so a vintage leaves the summation once y exceeds its build year plus lifetime and natural retirement is implicit in the summation window. $X_{y,i}$ contains the existing units whose operating year is at or before y and whose retirement year is after y. New, retrofit, and existing capacity can be retired early; extension and repower capacity cannot. The terms $\sigma^{\pm}$ are penalized slacks. An analogous expression tracks storage energy capacity.

3.2 Capacity pathways

Pathway	Specification
New build	Greenfield vintage, economically retireable, with its own book life
Life extension	Continuation of an existing vintage at end of natural life, offered only at the retirement boundary
Repower	Rebuild following retirement, admitted within a specified grace window
Retrofit	Conversion of retiring stock of another technology, tied to that technology's retirements by ratio groups

Extension is continuous at the retirement boundary; repower follows retirement within a lag window. Retrofit converts existing stock and adds no net capacity.

3.3 Storage duration

Storage energy capacity is bounded relative to power capacity by maximum and minimum duration limits, specified separately for new build, retrofit, and existing capacity. The existing-capacity minimum carries a penalized slack.

3.4 Resource potential and build-rate limits

Build is limited on four dimensions, each specified for technology groups at either global or zonal scope:

- Cumulative maxima, limiting capacity built through each year, representing resource potential.
- Cumulative minima, enforcing specified deployment.
- Annual maxima, limiting capacity built within a single year, representing industry build rate. The annual maximum may grow with the preceding year's build at a specified compound rate, and exceedance carries a penalty. An optional symmetric decline limit bounds the fall in a group's annual build relative to the preceding year's build.
- Annual minima.

Growth penalties apply a cost where annual capacity growth for a group exceeds a specified limit, providing a soft alternative to a hard annual bound. Land-use constraints, in primary and secondary forms, limit the aggregate area occupied by supply technology capacity, commodity supply, and transmission within each land-use zone group.

3.5 Portfolio and retrofit constraints

Technology ratio constraints enforce specified ratios between the capacities of technology groups within zones and years. Retrofit constraints tie retrofitted capacity to the retiring capacity of the originating technology and restrict retrofits to defined technology pairs.

4 Operations

4.1 Electricity balance

For each zone, representative day, and hour, generation plus net transmission inflow, less storage charging and over-generation, equals load grossed for losses:

$$\sum_{i} \mathrm{Gen}_{y,i,s,t} + F^{\mathrm{net}}_{y,z,s,t} - \sum_{k} \mathrm{Chg}_{y,k,s,t} - \mathrm{OG}_{y,z,s,t} \;=\; \left(1+\ell^{\mathrm{up}}\right)\left(L^{\mathrm{exo}} + L^{\mathrm{endo}} + L^{\mathrm{meas}} + L^{\mathrm{shift}} + L^{\mathrm{conv}}\right) - \mathrm{USE}_{y,z,s,t} \quad (2)$$

L^{exo} is exogenous load from EP, grossed for sector distribution losses (21); L^{endo} is the change in load from optimized demand technologies (31); L^{meas} is the adjustment from efficiency and fuel-switching measures (23); L^{shift} is the net effect of load shifting (Section 8.2); L^{conv} is electricity consumed by conversion technologies; and ℓ^{up} is an upstream loss factor. In mode (a), L^{endo} and L^{meas} are zero. Generation enters net of station use and transmission losses. Exports leave as flows to the trade zone within F^{net}. Unserved energy USE and over-generation OG are penalized slacks. Curtailment is not a balance term; it is derived as available less delivered output.

An optional hourly clean-energy matching constraint requires designated clean resources to cover a specified fraction of load within each hour over a cap-zone geography. The matched load comprises storage charging and conversion load. Primary and secondary matching requirements are specified separately.

4.2 Thermal generation

Unit commitment is specified per technology. For a technology without unit commitment, generation is bounded above by available capacity, times an hourly maximum profile where one is supplied, and below by an hourly minimum profile where one is supplied, and output is continuous between these bounds. Inter-hour ramp limits apply, and technologies

with zero ramp capability run at constant output within each representative day. Maximum and minimum annual capacity factors are specified by technology. The minimum load factor, start and shutdown costs, and minimum up and down times apply only under unit commitment.

Where unit commitment is represented, commitment is a continuous relaxation: committed capacity, startup, and shutdown are continuous variables, and committed megawatts are the committed unit count times a standard unit size. Generation is bounded between the minimum load factor and committed capacity. A commitment balance tracks committed capacity between hours as prior commitment plus starts less stops. Minimum up and down times are enforced without integer timers through two running stocks: capacity online long enough to be eligible to stop, and capacity offline long enough to be eligible to start. Decommitment is bounded by the stoppable stock and startup by the restartable stock; the stoppable stock is bounded by committed capacity and the restartable stock by offline capacity. Ramp bounds distinguish units continuing in operation from those crossing a start or stop boundary, with separate bounds across the inter-hour boundary. Startup and shutdown costs enter the objective.

4.3 Fixed-profile generation

Hourly generation from non-dispatchable technologies is bounded by available capacity times a time-varying profile factor. Curtailment is the difference between that product and delivered generation. Where negative generation is admitted, curtailment is separately bounded by potential generation.

4.4 Hydro generation

Hourly generation is bounded above by available capacity and below by a minimum flow requirement where specified. Ramp rates are bounded between consecutive hours, with optional multi-hour ramp limits. Energy is budgeted under one of two specifications: a cumulative energy constraint tracking generation within each period block against maximum and minimum limits, or a fixed daily energy budget per representative day. Under the cumulative specification, energy may be advanced or delayed within a specified allowance. Under fixed daily budgets, each representative day's generation equals its budget less penalized spill.

4.5 Electric storage

State of charge is tracked at three nested timescales so that long-duration storage does not require an hourly representation of the full year. The hourly level SOC^{H} tracks charge within a representative day and is cyclic except for long-duration technologies. The period level SOC^{D} is a running seasonal reservoir incremented by each representative day's net fill or drain, weighted by the calendar days that period represents. A window level bounds the seasonal envelope. With $\mathrm{SOC}^{H,\max}_{y,k,s}$ the peak intra-day state of charge on representative day s and $B_{y,k}$ the energy band reserved for planning-reserve pledges (Section 4.6), the physical bound is the chain

$$\mathrm{SOC}^{H}_{y,k,s,t} \le \mathrm{SOC}^{H,\max}_{y,k,s}, \qquad \mathrm{SOC}^{H,\max}_{y,k,s} + \mathrm{SOC}^{D}_{y,k,d} \;\le\; \mathrm{Window}_{y,k,\omega} \;\le\; \mathrm{Storage}_{y,k} - B_{y,k} \tag{3}$$

A further bound applies to the last calendar day of each period block. Hourly tracking accounts for charge, discharge, one-way efficiency, and self-discharge. Power throughput is bounded by rated power: discharge plus any planning-reserve pledge plus charge scaled by the charge-to-discharge ratio may not exceed power capacity. A piecewise penalty discourages deep discharge. Storage carries energy between representative days and is not tracked across a full chronological year. Where unit commitment is represented, storage charge and discharge carry ramp limits.

4.6 Planning reserves

Dependable capacity must cover a reserve margin above load in every hour of the reliability-critical representative days of Section 2.1:

$$(1+\mathrm{RM}_{z,y})\,\tilde{L}_{y,z,s,t} - \mathrm{UCR}_{y,z,s,t} \;\le\; \sum_{i} \mathrm{Cred}_{y,i,s,t} \tag{4}$$

$\tilde{L}$ is demand grossed for upstream losses, comprising exogenous and endogenous electric demand adjusted for measures and load shifting plus conversion electric load; RM is the reserve margin; UCR is a slack penalized at half the value of lost load. Credited supply applies a dependability factor, an effective-load-carrying-capability credit, under one of two crediting types specified per technology:

- Deemed resources contribute installed capacity times dependability.
- Performance resources contribute dispatch times dependability. Fixed-profile technologies contribute capacity times the hour's profile factor times dependability, so credit reflects hourly availability.

Performance-type transmission connections credit dependability times received flow, and deemed-type connections credit full capacity. Outflow is debited at dependability times flow grossed for line losses.

Storage contributes dependability times discharge less charging. Long-duration storage may also pledge reserve PR beyond its dispatch. A pledge is backed by an energy band $B_{y,k}$ that is withheld from cycling in (3). In each capacity-reservation window ω, dependability times the band must cover the pledged energy summed over every calendar day in the window:

$$\phi_{y,k}\,B_{y,k} \;\ge\; \eta_k^{-1} \sum_{s} n_{\omega,s} \sum_{t} \mathrm{PR}_{y,k,s,t} \tag{5}$$

where ϕ is dependability, η_k^{-1} the reciprocal of one-way efficiency, and $n_{\omega,s}$ the number of calendar days in window ω mapped to representative day s. Each calendar day counts once, independent of the clustering weights. The pledge also competes with dispatch and charging for rated power (Section 4.5), so the same megawatt is not credited twice.

For deemed conversion technologies an envelope variable bounds actual conversion electric load from above, supplying the demand-side analogue of deemed crediting.

4.7 Operating reserves

Where specified, upward and downward operating reserve requirements are met in each hour from online thermal, hydro, and storage capacity, with upward reserves drawn from headroom and downward reserves from footroom including storage charging.

4.8 Transmission

Electricity flow on each connection is bounded by available capacity hourly; carrier flows are bounded at representative-day or annual resolution with capacity scaled by the corresponding number of hours. Capacity is existing scheduled capacity plus cumulative new build, with incremental build equalized in both directions for bidirectional carriers. Cumulative build is bounded by corridor potential and annual build by a rate limit. Where a corridor is offered in cost tranches, tranches are built in order.

Hurdle-rate costs represent economic barriers to flow, specified either as a flat rate or as piecewise tranches in utilization. Under the piecewise specification, per-tranche flows sum to total flow and each tranche is bounded by a specified share of capacity, so successively larger flows incur successively higher rates. This is the mechanism applied to gas pipeline tariffs in Section 6.8.

Net flow terms enter the electricity balance (2) and the carrier balances (6). Annual transmission losses are accumulated per carrier and zone.

4.9 Network peak and distribution cost attribution

Distribution and local transmission capacity is not represented as buildable. Its cost is attributed to served peak load. System peak per zone and sector peak per distribution feeder are decision variables bounded below by net load in every hour of the reliability-critical representative days. A positive per-unit cost in the objective drives each to the maximum co-optimized load. A monotonic non-decrease condition and a minimum floor prevent the imputed peak from falling below observed load or declining as though infrastructure were removed. Because the peak is derived from optimized dispatch, storage, load shifting, demand-side technology choice, and trade reduce imputed network cost directly. The peak constraint is distinct from the planning reserve of Section 4.6: it applies no reserve margin and uses actual dispatch, sharing only the reliability-day index set.

5 Sector coupling

5.1 Carrier balances

Each carrier is balanced either physically at a specified resolution or through an annual accounting identity. Physical balances admit storage chronology and sub-annual transport; accounting balances apply to quantities settled annually without obligation of physical delivery, including credit and book-and-claim instruments.

The physical balance, at the resolution tagged for carrier β, requires supply to equal demand in every period:

$$\begin{aligned} &\mathrm{Dis}^{\beta} + \mathrm{ComIn}^{\beta} + \mathrm{ConvOut}^{\beta} + F^{\mathrm{net},\beta} + \mathrm{Uns}^{\beta} \\ &\quad = \mathrm{Chg}^{\beta} + \mathrm{ComOut}^{\beta} + \mathrm{ConvIn}^{\beta} + \mathrm{Burn}^{\beta} + D^{\mathrm{exo},\beta} + D^{\mathrm{meas},\beta} + D^{\mathrm{endo},\beta} + \mathrm{Ovs}^{\beta} + \mathrm{Spl}^{\beta} \end{aligned} \tag{6}$$

Supply terms are storage discharge, commodity inflow, conversion output, and net transport inflow; demand terms are storage charge, commodity outflow, efficiency-weighted conversion input, generator fuel consumption, exogenous demand (22), the measure adjustment (23), and endogenous demand from optimized demand technologies (Section 8.9), all grossed for upstream losses. In mode (a), D^{meas} and D^{endo} are zero. Unserved and overserved quantities are penalized slacks. For carriers tagged as surplus-tolerant, a spill term is subtracted from supply. Terms at finer resolution are prorated into coarser balances.

The accounting balance requires annual supply to equal annual demand, with a separate annual exceedance slack for carriers allowed to be over-supplied, and with no sub-annual index and no storage state. A physically balanced carrier additionally receives the annual identity, reconciled to the physical solution by a penalized slack.

5.2 Carrier storage and transport

Carrier storage is bounded in injection and withdrawal by power capacity and in state of charge by energy capacity, at hourly or representative-day resolution as tagged. State of charge is tracked within and across period blocks. Annual storage losses are accumulated. A piecewise penalty applies to low state of charge for hourly carrier storage. Carrier transport uses the transmission formulation of Section 4.8 at the carrier's tagged resolution.

5.3 Conversion technologies

Conversion technologies link carriers and electricity. Output is bounded by capacity at the tagged resolution and by annual maximum throughput and minimum capacity factor. Technologies with modal efficiency operate across discrete modes with distinct efficiencies and co-product output ratios. Per-mode activity sums to total activity, and each mode's activity is bounded by a specified share of the total, at annual, representative-day, and hourly resolution. Mode output ratios fix the relative quantities of co-products. Electricity consumption is computed at the tagged resolution, with per-mode consumption folded into the same expression. Maximum carrier throughput limits may be imposed on designated technology subgroups at hourly and representative-day resolution.

The conversion technologies represented include water electrolysis, steam methane reforming, autothermal reforming with capture, methanation, Fischer–Tropsch synthesis, biomass gasification to gases and liquids with capture, fast pyrolysis with capture, cellulosic and corn ethanol, hydroprocessed esters and fatty acids, fatty acid methyl ester biodiesel, alcohol-to-fuels, anaerobic digestion and biogas upgrading, hydrogen and natural gas liquefaction, hydrocarbon gas liquid fractionation, oil and gas extraction and processing (Section 6.3), petroleum refining, industrial steam and heat supply (boilers on electricity, hydrogen, pipeline gas and legacy fuels, heat pumps, and thermal storage), and the industrial material routes of Section 5.4.

5.4 Commodity supply and industrial materials

Commodity supply is carried by an allocation variable from supplying to consuming zone and a per-bin total. The annual and cumulative potential limits are

$$\frac{1}{W} P_{y,c,b,z} \le \bar{P}_{y,c,b,z}, \qquad \frac{1}{W} \sum_{y' \le y} g_{y'} P_{y',c,b,z} \le \Phi_{c,b,z} + \sigma \tag{7}$$

with $\bar{P}$ annual potential, Φ cumulative potential, and σ a penalized slack. Four further constraints apply. A non-decrease constraint requires designated commodity allocations to be no lower than in the preceding year, with penalized slack. A bin-to-total identity ties per-bin totals to the sum of zonal allocations. A trajectory constraint sets a floor on each bin's production relative to the preceding year, $\sum P_y \ge \tau^{g_y} \sum P_{y-1}$, which caps the rate of decline. A joint constraint keeps the trajectory commitment and the cumulative limit mutually feasible by requiring that past cumulative production plus production carried forward by the trajectory not exceed cumulative potential. Commodity cost is accumulated into the objective. Primary resources represented comprise biomass feedstocks, oil and gas (Section 6), coal, and CO_2 sequestration capacity.

Steel, clinker and cement, lime, and ammonia are represented as products with their own production technologies and process emissions, competing on the same basis as energy supply routes. Represented routes include blast furnace–basic oxygen furnace with and without capture, coke production with and without capture, direct reduced iron on natural gas with and without capture and on hydrogen, electric arc furnace, downstream rolling and casting (including hydrogen-fired), conventional clinker with kiln burners on natural gas, hydrogen, biomass and municipal solid waste, clinker with direct-separation capture (new build and retrofit) and oxyfuel capture, limestone calcined clay cement as a clinker substitute, conventional and capture-equipped lime, and Haber–Bosch ammonia synthesis.

Upper and lower limits on the share of each carrier supplied by individual commodities, conversion technologies, and groups of conversion technologies apply by zone and year. Delivery cost is accumulated per carrier and zone.

5.5 Emissions accounting

Emissions are tracked per gas g and zone. Total emissions sum four terms:

$$E_{y,z,\mathrm{g}} \;=\; E^{\mathrm{com}}_{y,z,\mathrm{g}} + E^{\mathrm{del}}_{y,z,\mathrm{g}} + E^{\mathrm{conv}}_{y,z,\mathrm{g}} + E^{\mathrm{tx}}_{y,z,\mathrm{g}} \tag{8}$$

comprising commodity consumption, carrier delivery, conversion technology operation, and electricity transmission, the last computed as flow times an exogenous emission rate per connection. Gases are aggregated by global-warming-potential weighting specified per zone and year. A gas with zero weight is tracked and reported but does not count toward any cap, price, or standard.

The emissions instruments are annual caps on total emissions over primary and secondary zone groups, a cumulative cap over the horizon, a per-gas per-zone emissions price, a radiative-forcing limit over specified warming years, and the sectoral carbon-intensity and absolute emissions standards of Section 7. Each instrument is switched on in the run configuration and is active only where its input table carries finite values.

Clean-energy standards are formulated separately, as primary and secondary standards, each requiring clean generation to meet a specified share of load over its own enforcement geography. Generation and demand are tracked in the model zones and mapped to the enforcement geography. Qualifying output is counted from supply-side generation, from demand-side generation and conversion, and from qualifying thermal generation through the fuel it burns. Renewable energy credits are tracked, with limits on credits claimed from out-of-zone generation; a delivered-energy option requires such generation to be matched by import flows. Incentive payments on qualifying generation and price caps are optional.

6 Hydrocarbon supply representation

Domestic oil and gas supply is endogenous: quantity, price, and product composition are determined by the optimization. The supply representation is parameterized by play inventories, per-well production profiles, drilling and facility costs, dry-hole rates, royalty and severance rates, technology improvement rates, rig and footage regressions, and drilling-pace parameters, all of which are input data. Three quantities are calibrated to a reference drilling projection: annual potential in the first optimization year, set to reference drilling in that year scaled by a headroom factor for gas and for oil; the first-year rig-capacity floor; and the drilling-schedule decline of Section 6.4, which is linearized along the reference cumulative drilling path.

6.1 Resource discretization

Plays. Shale gas and tight oil are disaggregated into named plays, each with a residual play for resource outside the named plays. Conventional gas, conventional oil, and coalbed methane are represented at type level. Offshore and Alaska oil and gas and Canadian gas enter as separate commodities with exogenous annual potential. Each play is a distinct commodity with its own costs, potential, and drilling parameters. Plays of a given type supply a single input carrier and are processed by one type-level extraction-and-processing technology. Results are reported at type level.

Bins. Within each producing district and commodity, play slots are ordered by breakeven cost and grouped into bins on fixed breakeven-cost edges, specified separately for gas (per unit energy) and oil (per barrel), with a cost ceiling above which bins are excluded. Bin indices are then assigned globally per commodity, one index per distinct cost level, ascending in cost, across all districts. Because the model zone geography is coarser than the district geography of the input data, global cost-keyed indexing yields exact aggregate costs under district-to-zone aggregation. A bin's cost is the resource-weighted average of its play slots. It is uniform within the bin unless intra-bin escalation is specified (Section 6.5); resource exhaustion is governed by the cumulative limit of (7) and the drilling constraints of Section 6.4.

Potential and decline. Each (commodity, bin, zone) carries cumulative potential $\Phi_{c,b,z}$, annual potential $\bar{P}_{y,c,b,z}$, and trajectory parameter $\tau_{y,c,b,z}$, the fraction of one year's production persisting to the next absent new drilling. These enter the commodity constraints of Section 5.4.

Legacy production. Production from wells drilled before the horizon is represented by separate legacy commodities with declining annual potential and zero marginal resource cost. Legacy and new supply of a given type supply the same carrier and are processed by the same technology.

6.2 Breakeven supply cost

The cost assigned to a play slot is the price at which the after-tax net present value of drilling one well is zero, computed from a project cash flow. Let K denote capital charged at $t = 0$, comprising drilling cost, facility capital, and probability-weighted expected dry-hole cost per well drilled; O_t annual operating cost, comprising per-unit production and transport cost, per-unit production severance tax, and a fixed share of general and administrative cost; Q_t annual production; ρ the

revenue retention factor net of royalty and revenue-based severance tax; δ_t the discount factor at the project hurdle rate; T the combined state and federal income tax rate with state tax deductible; D the present value of capital tax deductions; and f_t a forward price index. Then

$$p^* = \frac{K - TD + (1-T)\sum_t \delta_t O_t}{(1-T)\sum_t \delta_t f_t \rho Q_t} \tag{9}$$

The hurdle rate is a project discount rate, computed as a real weighted average cost of capital plus an extra-risk premium and a required return. Intangible drilling cost and dry holes are expensed at $t = 0$, and the tangible share is depreciated on a fixed schedule. Severance and production taxes are state-specific, with revenue-based rates applied to post-royalty revenue and per-unit taxes treated as operating cost. The forward price index renders p^* a current-year-equivalent breakeven under a rising price path.

Breakevens decline over the horizon at technology-improvement rates applied separately to drilling and dry-hole cost, to facility and operating cost, and to per-well production, with higher rates for emerging plays in early years. Bin-level cost multipliers are resource-weighted from the constituent slots.

6.3 Extraction and processing technologies

Extraction and processing are represented as one technology within the modal conversion formulation of Section 5.3. Its modal efficiency coefficients specify the wellhead product split and processing energy consumption:

- Gas plays yield dry pipeline gas and hydrocarbon gas liquids in district-specific proportions, with a share of the gas stream consumed as field and compression fuel.
- Oil plays yield crude, associated gas, and natural gas liquids in proportions set by the district gas-to-oil ratio.
- Coalbed methane yields dry gas.

Three variants are defined per type: a base variant consuming produced gas as field fuel; an electrified variant consuming electricity in place of that gas, at a lower energy input that reflects motor efficiency relative to engine efficiency; and a carbon-capture variant applied to the acid-gas removal stream. Capture variants are undefined for coalbed methane, which has no acid-gas stream.

Each variant operates in two modes distinguishing domestic from export markets. Hydrocarbon gas liquids are fractionated into components by a separate conversion technology with its own midstream costs.

The modal output split and its carbon weights. The modal coefficients are what make the co-product attribution of Section 7.3 operate on the hydrocarbon chain, so the split is set out here explicitly. For one unit of raw wellhead energy entering a gas extraction and processing technology, the modal output shares over pipeline gas, hydrocarbon gas liquids, and crude sum to one, and a further share of the pipeline gas stream is consumed as field and compression fuel. The shares are specific to each model zone, aggregated from the producing districts it contains, and reflect local gas wetness. Each output carrier has its own carbon content per unit energy, and these contents weight the co-products. Pipeline gas has the lowest content, hydrocarbon gas liquids an intermediate one, and crude the highest. The contents are the same everywhere, being set by the carbon chemistry of the product. Geographic variation enters through the shares: in wetter zones the energy share of pipeline gas is lower and the balance moves to liquids and crude.

Two consequences follow. First, the wellhead carbon charge is the share-weighted sum of the output contents, so it varies with composition and does not take a single dry-gas value. Each commodity's charge is computed from the output split of its producing district, and it exceeds the content of pure methane by an amount that rises with the liquids and crude share. The commodity charges and the zone-level processing slates are aggregated from district data separately, so they agree at district level but not exactly at zone level. The same composition weighting is applied to legacy production, which pools into the same carriers and is processed by the same technology. Second, the carbon-weighted shares differ systematically

from the energy shares, and by more in wetter districts: hydrocarbon gas liquids and crude take a larger share of the carbon than of the energy, and pipeline gas a smaller one. That divergence is the reason the numerator of a product-group constraint apportions by carbon while the denominator apportions by energy (Sections 7.3 and 7.4).

6.4 Drilling capacity constraints

Production is constrained by drilling capacity through a constraint on new extraction. New extraction is production added by new drilling in excess of that delivered by the declining existing stock:

$$N_{y,c,b,z} \;\geq\; (1-\tau_{y,c,b,z})\,P_{y,c,b,z} \;+\; \tau_{y,c,b,z}\,\frac{P_{y,c,b,z}-P_{y-1,c,b,z}}{g_y}, \qquad N_{y,c,b,z} \;\geq\; 0 \tag{10}$$

N carries no direct cost. Equation (10) sets its floor and the constraints below cap or price it, so at the optimum N sits at the floor without integer variables. The ramp constraint (16) is the exception: a higher N_{y-1} relaxes it, so N can exceed its floor where the ramp binds.

Equation (10) follows from the cohort identity $P_t = D_t + \tau P_{t-1}$ for annual drilling D_t, with P_{y-1} interpolated between optimization years. At steady state it reduces to $(1-\tau)P$. It is the tangent plane at $P_y = P_{y-1}$ of the corresponding geometric expression, exact at steady state, and it understates the drilling requirement by 1 to 5 percent under a 1.5- to 3-fold expansion.

The zero floor on N is applied per bin and zone before aggregation, so a declining bin does not offset the drilling requirement of a growing bin within a resource group. The first optimization year carries no new-extraction variable and is exempt from every drilling constraint of Sections 6.4, 6.6, and 6.7.

For each year, group, and constraint region:

$$\frac{1}{W}\sum_{(c,b)\,\in\,\mathcal{M}(\mathrm{grp})}\;\sum_{(z,\,\theta)\,\in\,\mathcal{G}(\mathrm{grp},\,\mathrm{gau})}\theta\cdot N_{y,c,b,z}\cdot\kappa_{\mathrm{grp},c,b} \;\leq\; C_{\mathrm{grp},\,\mathrm{gau},\,y} \tag{11}$$

$\mathcal{M}$ is the group member set of (commodity, bin) pairs; $\mathcal{G}$ maps model zones to the group constraint geography with allocation ratio θ; κ is a contribution factor in resource units per unit energy; C is the physical ceiling. Units resolve as (energy yr^{-1})(resource energy^{-1}) = resource yr^{-1}. Each group specifies its own constraint geography.

Two group families are defined.

Rig capacity, in wells per year, one group for each drilling region plus a national group, with ceilings from a rig-capacity regression on the real oil price. Gas and oil draw on a common fleet. Regional ceilings carry a slack allowance for interregional rig relocation; the national ceiling is the sum of pre-slack regional values. Dry holes are included in the well count. Drilling to maintain committed producing projects is netted from available capacity, because that production is carried in the legacy commodities. Contribution factors are weighted by drilling capacity. Ceilings are expressed in productivity-scaled well-equivalents anchored to the base year, because per-well productivity changes over the horizon while the regression is calibrated on a fixed basis.

Footage capacity, in feet per year, one group per drilling region, with ceilings from a footage regression. Member contributions are the rig contribution scaled by average footage per well.

Drilling-schedule decline. A play's drilling rate holds at its plateau until cumulative drilling passes a spacing threshold in wells, after which it decays geometrically in time. Geometric decay in time is linear in cumulative drilling, so the decay is imposed as

$$\frac{1}{W}N_{y,c,b,z} \;+\; \lambda_{y,c,b,z}\,\frac{1}{W}\sum_{y'\leq y} g_{y'}N_{y',c,b,z} \;\leq\; \bar{D}_{y,c,b,z} \tag{12}$$

with plateau $\bar{D}$ and slope λ aggregated over the play's projects, each project declining against its own threshold. Because the aggregate is piecewise linear, $\bar{D}$ and λ are re-linearized in each year at the reference cumulative drilling. The constraint limits the drilling rate, not remaining resource.

Per-well recovery is independent of spacing: the drilling schedule constrains the drilling rate beyond a spacing threshold, and no parent–child interference term reduces per-well production.

6.5 Intra-bin cost escalation

A bin aggregates multiple play slots, and its assigned cost is a resource-weighted average. Intra-bin cost escalation represents the cost dispersion within a bin. The bin's cumulative potential is divided into $K_{c,b}$ equal-resource depletion tranches k, each carrying a non-negative cost adder that is non-decreasing in k. Bin production is allocated across the tranches, and each tranche can supply no more than its share of cumulative potential:

$$P_{y,c,b,z} \;=\; \sum_{k} S_{y,c,b,z,k}, \qquad \frac{1}{W}\sum_{y'\le y} g_{y'}\, S_{y',c,b,z,k} \;\le\; \frac{\Phi_{c,b,z}}{K_{c,b}} + \sigma \tag{13}$$

where S is tranche production and σ a penalized slack. Because the adders are non-decreasing in k, cost minimization fills tranches in index order and no ordering constraint is required. The adders enter the objective as the intra-bin depletion cost of Section 2.3. A bin without tranche data carries no adder.

6.6 Drilling-pace price response

Wellhead price is set by the cost of the marginal well plus the price required to sustain a rate of drilling. Each play is drilled at a plateau rate scaled by $(p/P_B)^{1/s}$ below the base price P_B and by $\min\big((p/P_B)^{s},\,\Lambda\big)$ above it, with s a play-specific sensitivity exponent and $\Lambda = 2$ a surge ceiling. Inverting, a pace fraction φ of plateau requires price $P_B\varphi^{e(\varphi)}$, with $e = s$ for $\varphi \le 1$ and $e = 1/s$ for $\varphi > 1$.

New drilling per play, measured as the first-year output of new wells, is divided into N_{pace} equal-width tranches spanning zero to Λ times plateau. Tranche k carries reservation premium

$$\pi_{y,c,b,k} \;=\; \frac{\max\!\Big(0,\;\; P_B\,\varphi_k^{\,e(\varphi_k)} \;-\; c_{y,c,b}\Big)}{\max\big(1-\tau_{c,b},\;\underline{\tau}\big)}, \qquad \varphi_k \;=\; \Big(k-\tfrac{1}{2}\Big)\frac{\Lambda}{N_{\text{pace}}} \tag{14}$$

where $c_{y,c,b}$ is bin cost in year y and φ_k is the tranche midpoint. The denominator converts the premium from a per-unit-of-production to a per-unit-of-new-drilling basis; it is floored at $\underline{\tau}$ for bins with high persistence τ. The premium is evaluated against the technology-declined cost of the current year, so technology improvement is not applied twice.

P_B is a base gas price for gas and a per-play base oil price for oil, expressed on the model energy basis through the denominator of (9). The exponent s depends on well type: it is a function of per-well productivity for shale gas and tight oil and is unity for conventional resources and coalbed methane, clipped to a specified range. For oil wells with a high first-year gas-to-oil ratio the exponent is further scaled by the ratio of the gas price to the base gas price.

Premium tranches are indexed by (commodity, bin, tranche) without a zone index. The linking constraint sums new extraction over zones:

$$\sum_{z} N_{y,c,b,z} \;=\; \sum_{k} R_{y,c,b,k}, \qquad \frac{1}{W}R_{y,c,b,k} \;\le\; \frac{1}{N_{\text{pace}}}\sum_{z} \Lambda\,\bar{D}^{\text{pace}}_{y,c,b,z} \tag{15}$$

where $\bar{D}^{\text{pace}}$ is the drilling plateau. Drilling pace is a property of the play, not of a consuming zone. Indexing by zone would subject a demand-heavy zone's share of a play to surge pricing while the play as a whole is below plateau. Premiums are input data and enter the objective as a separate component. The dual value of the linking constraint is the

play's marginal pace rent.

6.7 Project ramp constraint

Development of a play approaches plateau over a multi-year ramp:

$$N_{y,c,b,z} - N_{y-1,c,b,z} \;\le\; g_y \cdot W \cdot \frac{\bar{D}^{\text{pace}}_{y,c,b,z}}{Y_{\text{ramp}}} \tag{16}$$

Y_{ramp} is a ramp period of five years and $\bar{D}^{\text{pace}}$ the drilling plateau of (15), so a play reaches plateau from zero in five years. Where a bin has no plateau data, the plateau is approximated by $(1-\tau)\bar{P}/\Lambda$.

The constraint is anchored to the year a bin is first developed, so inventory developed later in the horizon ramps from zero at that point. In the first constrained year the right-hand side is seeded with the first-year annual potential, so plays already drilling continue from that pace.

6.8 Midstream, trade, and upstream emissions

Pipelines. Interregional gas pipelines are represented as a network with expandable capacity under Section 4.8. Transport is priced by piecewise tranches in utilization, approximating a cost-of-service tariff at normal utilization and a scarcity premium approaching capacity.

Trade. Pipeline gas and petroleum coke exports enter as exogenous demand. Liquefied natural gas exports face a price-responsive export curve with constant elasticity, offered in bins up to a specified multiple of a reference export volume, and liquefaction, including electric-drive liquefaction, is buildable capacity. Crude oil and hydrocarbon gas liquid exports are also price-responsive. Refined product exports are fixed at a reference volume. Unfinished oil, refined product, and blending component imports are available subject to annual limits.

Wellhead CO_2. Carbon charged at the wellhead equals the carbon content of the output slate of the producing district. Slate composition varies by district and its components differ in carbon content per unit energy, dry gas lowest and crude and condensate highest; wellhead rates are therefore derived from the district split, for legacy and new supply alike (Section 6.3).

Fugitive methane. Upstream methane rates for new onshore supply are resolved by producing district from play-level emission factors, scaled to preserve the production-weighted national mean. Legacy, Alaska, offshore, and Canadian supply use type-level national rates. Methane enters the emissions accounting of Section 5.5 through the global-warming-potential weights.

Process CO_2. CO_2 removed from the raw gas stream during processing is an output of the processing technology and is not a combustion product.

7 Sectoral carbon-intensity standards

A sectoral carbon-intensity standard constrains emissions per unit of a specified product. Representation requires a system boundary per product, an accounting rule for carbon crossing that boundary, a denominator, and a co-product attribution rule. Standards are formulated as constraints on product groups: sets of energy carriers or material products, aggregated over a set of zones, constrained in absolute emissions or in emissions intensity. Group memberships, ceilings, and cap zones are input data.

7.1 Product group definition

Each carrier declares membership of a group's input set (IN) and of a group's product set (OUT). Three cases follow.

A carrier in IN but not OUT is an intermediate. Its carbon is charged where it enters the boundary and it is not a group product; it enters neither numerator nor denominator.

A carrier in both IN and OUT is group product consumed by the group. Its volume enters the denominator, and it is charged in the numerator at its deemed intensity over gross supply. This treatment is equivalent to netting own use from the denominator and does not require identification of own-use flows.

A carrier in OUT but not IN is group product not consumed within the boundary; it enters the denominator only.

7.2 Constraint formulation

For year y, group γ, and cap zone ζ, let E denote attributed emissions and S product supply. The absolute constraint is

$$E_{y,\gamma,\zeta} \;\le\; \bar{E}_{y,\gamma,\zeta} \tag{17}$$

with $\bar{E}$ the emissions ceiling. The absolute form carries no trading term.

The intensity constraint is imposed in cross-multiplied form,

$$E_{y,\gamma,\zeta} \;\le\; \bar{\iota}_{y,\gamma,\zeta}\, S_{y,\gamma,\zeta} \;+\; \theta_{y,\gamma,\zeta} \tag{18}$$

with $\bar{\iota}$ the intensity ceiling and θ the trading term (Section 7.5). Because $\bar{\iota}$ is a parameter, cross-multiplication of $E/S \le \bar{\iota}$ preserves linearity. The denominator remains endogenous: compliance is achieved by reducing emissions or by increasing compliant supply.

An infinite ceiling removes the constraint for that group, zone, and year. The emissions and supply accounting variables remain defined, so measured intensity is reported for unconstrained groups on the same basis used to enforce a ceiling. Cap zones may be coarser than model zones: a membership map assigns model zones to each cap zone, and group emissions and supply are summed over the member zones before the ceiling is applied.

Equation (17) excludes all IN carriers from its numerator, having no denominator subject to dilution. Equation (18) excludes intermediates only and applies the treatment of Section 7.1.

7.3 Emissions attribution

The numerator sums four categories of carbon entering the group boundary. Processes producing any carrier inside the boundary, intermediates included, are charged for the carbon in their inputs; inputs that are themselves intermediates of the group are not charged again.

Modal conversion technologies. A mode co-producing multiple outputs from a shared feedstock attributes feedstock carbon among co-products by carbon-weighted output share:

$$E^{\text{modal}} = \sum_{(i,n)} \Big(\sum_{o \in \text{IN}(\gamma)} w^{\text{C}}_{y,i,n,o} \Big) \sum_{\beta \notin \text{IN}(\gamma) \setminus \text{OUT}(\gamma)} X_{y,i,n}\, \eta^{\text{in}}_{i,n,\beta}\, \iota^{\text{d}}_{z(i),y,\beta} \tag{19}$$

X is mode activity; η^{in} its input intensity for carrier β; ι^{d} the deemed carbon intensity of that carrier in the technology's zone $z(i)$; w^{C} a carbon-weighted output share summing to unity across the mode's outputs, falling back to energy shares where carbon contents are not defined.

Non-modal conversion technologies contribute allocation times efficiency times deemed intensity. Deemed intensity is evaluated at the technology's own zone, where input consumption occurs; resulting emissions are booked to the zone receiving the output.

Commodity inflows contribute allocation times commodity emission rate times global warming potential per gas. The upstream emissions of Section 6.8 enter the standards through this term.

Generation contributes where electricity is an IN carrier.

7.4 Product supply accounting

The denominator sums the same four categories restricted to OUT carriers: modal conversion output, non-modal conversion allocation, commodity inflow allocated to OUT carriers, and, where electricity is an OUT carrier, generation weighted across representative days. The modal term uses an energy-weighted output share $w^{\mathrm{E}}_{y,i,n,o}$, in contrast to the carbon-weighted share in (19); both sum to unity across a mode's outputs.

7.5 Deemed emission intensities and trading

Numerator charges are priced through deemed carbon intensity ι^{d}, indexed by carrier, zone, and year. A carrier's deemed intensity may be specified as its direct combustion carbon or as a lifecycle value including upstream production and processing emissions; applying identical ceilings on alternative bases isolates the effect of the accounting boundary from that of the physical system. Deemed intensity is specified for each carrier in every zone where it bears carbon. Carriers without input rows, including biogenic feedstocks, ores, and uranium, carry a deemed intensity of zero.

By default, where a group carries a finite ceiling, the deemed intensity of each of its product carriers is set to that ceiling, by zone and year. A product consumed by another group is then charged at its compliant intensity instead of its input value. Two cases are treated separately: a deemed intensity of zero in the input is retained, and in years with an infinite ceiling the product carriers are deemed at zero. A run option leaves the input deemed intensities in force instead of substituting the ceiling.

The trading term is zero-sum within a cap zone,

$$\sum_{\gamma} \theta_{y,\gamma,\zeta} \;=\; 0 \tag{20}$$

so a group exceeds its ceiling only to the extent another falls below its own, and the standards within a cap zone trade as one market. Groups with infinite ceilings have θ fixed at zero.

8 Demand representation and demand-side optimization

This section gives the demand terms that enter the balances in both modes (Sections 8.1 and 8.2), the formulation of demand-side optimization in mode (b) (Sections 8.3 to 8.9), and the variables and constraints specific to mode (b) (Section 8.10).

8.1 Exogenous demand

Exogenous demand is read from EP in both modes. For electricity, EP supplies hourly load $H^{\mathrm{EP}}_{y,s,t,z,w,j}$ by zone, demand sector w, and distribution feeder j, taken from the weather-year load series at the hours of each representative day. The exogenous term of the electricity balance (2) is

$$L^{\mathrm{exo}}_{y,z,s,t} \;=\; \sum_{(w,j)} H^{\mathrm{EP}}_{y,s,t,z,w,j}\,\bigl(1+\ell_{z,w,y}\bigr) \tag{21}$$

with $\ell_{z,w,y}$ the distribution loss factor of sector w. For every other carrier EP supplies annual demand $A_{z,y,\beta,w}$ by zone, year, and sector. A carrier balanced at hourly resolution also receives an hourly profile, and its demand enters (6) hour by hour; a carrier balanced annually enters at its annual total. The exogenous term of (6) is

$$D^{\mathrm{exo},\beta}_{y,z,s,t} \;=\; \sum_{w} A_{z,y,\beta,w}\,h_{y,s,t,z,\beta,w}\,\bigl(1+\ell_{z,w,y}\bigr)\;\;(\text{hourly}), \qquad D^{\mathrm{exo},\beta}_{y,z} \;=\; \sum_{w} A_{z,y,\beta,w}\,\bigl(1+\ell_{z,w,y}\bigr)\;\;(\text{annual}) \tag{22}$$

where h is the hourly profile normalized so that its representative-day-weighted sum over the year is one. The annual total of every carrier, including those balanced hourly, also enters the annual accounting identity of Section 5.1.

The exogenous quantities cover every demand subsector. EP computes them with its own technology adoption and with the efficiency and fuel-switching measures of the EP case applied. In mode (a) they are the whole of demand. In mode (b) they are the reference about which RIO adjusts: for an optimized subsector they are the demand of the EP baseline sales, and the endogenous term (31) adds the change in demand caused by departures from those sales.

Measure adoption enters as an adjustment to the exogenous terms. Let $M_{m,y,z} \in [0,1]$ be the fraction of measure m that RIO does not adopt, relative to its full adoption in the EP case, ς the measure's savings of a carrier, and a the additions of another carrier made by a fuel-switching measure. The measure term of the electricity balance is

$$L^{\text{meas}}_{y,z,s,t} \;=\; \sum_{m,(w,j)} M_{m,y,z}\left(\varsigma_{y,z,m} - a_{y,z,m}\right) \hat{H}_{y,s,t,z,w,j}\, \alpha^{\text{fdr}}_{z,y,m,j} \tag{23}$$

where $\hat{H}$ is the exogenous load shape of sector w and feeder j normalized to an annual sum of one and α^{fdr} the feeder allocation of the measure's subsector. The carrier term $D^{\text{meas},\beta}$ of (6) has the same form with the carrier profile h at hourly resolution, or the annual quantity at annual resolution, each grossed for sector losses. Not adopting a measure restores the savings it would have made and removes the additions it would have made. In mode (a), and whenever the measure switch is off, M is fixed at zero and the measures are adopted as in the EP case.

8.2 Load shifting

Load shifting is formulated separately from technology adoption. It bounds charge and discharge of a flexible load, per zone, feeder, and subsector, by a specified flexible fraction of that subsector's exogenous load profile, supplied as flexible-load input data. The shifted energy follows a state-of-charge recursion with advance and delay limits, under one of two regimes selected in the run configuration: shifting between days along the period chronology, or balancing within the representative day. Net shifted load enters (2) as L^{shift}, and load-shifting cost enters the objective.

The shifting bounds are defined on exogenous subsector load and carry no endogenous term, so load shifting applies to the load fixed by EP. Load shifting and demand technology optimization are not used together.

8.3 Decision scope in mode (b)

In mode (b), RIO determines demand-side technology sales by year and zone, within bounds supplied by EP, jointly with supply-side investment. Adoption of a technology is therefore determined against the cost of supplying the carriers it consumes and displaces and of the network capacity it requires. Adoption responds to the marginal cost of those carriers, evaluated at the dual values of the corresponding balance constraints, subject to the payback-acceptance allocation of Section 8.6. Substitution is represented at the level of individual technologies delivering a specified energy service, without an elasticity on aggregate demand.

EP supplies energy service demand and its drivers, equipment stock rollover and retirement, technology cost and efficiency by vintage, hourly profiles by technology, and baseline sales. RIO determines sales quantities within the EP-defined envelope; it does not re-derive service demand, advance stocks, or introduce technologies not characterized by EP. Each demand technology carries a single vintage, so sales variables are indexed sparsely by vintage year.

Optimization is designated per subsector in the input data, and the designation takes effect only when the demand technology switch is on. Demand technology data are read only for designated subsectors. A subsector that is not designated is handled as in mode (a): its demand is the exogenous demand of Section 8.1, and RIO can modify it only through the efficiency and fuel-switching measures of Section 8.8.

8.4 Adoption bounds

Sales bounds are formed from EP sales trajectories supplied as minimum, maximum, and baseline tables. For each demand group, technology, zone, and vintage, RIO sets the floor to the smaller and the ceiling to the larger of the minimum and maximum tables. When a single EP case supplies all three tables, sales are fixed at the baseline up to the penalized slack of Section 8.5.

The adoption envelope is therefore set by the choice of EP cases. Two EP cases that differ only in their efficiency, fuel-switching, and sales-share measures span the adoption range between them without confounding changes in other assumptions: the case with the measures supplies the baseline and the maximum, and the case without them supplies the minimum. Where the measures shift sales toward electric technologies, those technologies mostly have the baseline at the ceiling, so their adoption can be reduced, and the fossil technologies they displace mostly have the baseline at the floor, so their adoption can be increased. The conservation identity (25) couples the two.

8.5 Sales and service demand

Let $\mathrm{Sales}_{g,\tau}$ denote sales of technology τ in demand group g, m_g a group scaling multiplier, $\mathrm{Sales}^{\mathrm{EP}}$ the EP baseline, and ϕ a service-demand factor. Service demand is specified as a deviation from baseline:

$$\mathrm{SD}_{y,g,\tau} \;=\; \Big(\mathrm{Sales}_{g,\tau} - \frac{\mathrm{Sales}^{\mathrm{EP}}_{g,\tau}}{m_g}\Big)\,\phi_{y,g,\tau} \tag{24}$$

Baseline energy demand embodies EP's stock accounting and is carried in the exogenous terms of Section 8.1; RIO adjusts about it. Deviations within a group sum to zero, conserving service demand:

$$-\varepsilon \;\le\; \sum_{\tau \in g,z} \mathrm{SD}_{y,g,\tau} \;\le\; \varepsilon \tag{25}$$

imposed as a tolerance band that accommodates independent rounding and rescaling of the terms in (24).

Sales are bounded by the envelope of Section 8.4, with penalized slack on both bounds. Technologies subject to a linkage relation (Section 8.7) are excluded; the lower bound applies where positive.

8.6 Payback-acceptance allocation

Purchasers exhibit a distribution of implicit discount rates. Sales are allocated across payback-acceptance bins, with $\alpha_{g,p}$ the share of group g's market requiring payback within bin p:

$$\sum_{\tau \in g,z,y} \mathrm{Sales}_{g,\tau}\,\alpha_{g,p} \;=\; \sum_{\tau \in g,z,y} \mathrm{Sales}^{\mathrm{pb}}_{p,g,\tau}, \qquad \mathrm{Sales}_{g,\tau} = \sum_{p} \mathrm{Sales}^{\mathrm{pb}}_{p,g,\tau} \tag{26}$$

A technology satisfies the hurdle of the bin whose market share it obtains, so partial adoption is admitted without a single system-wide hurdle. The shares $\alpha_{g,p}$ form a payback-acceptance curve by demand group and are input data, normalized to sum to one within each group.

The bin index p is a real-valued acceptance threshold in years. Capital cost is charged to bin p as $C_{g,\tau}/p$, the year-one capital contribution, so a purchaser in bin p adopts when that annualized charge is covered by annual operating savings, which is the condition that simple payback not exceed p.

Relation of thresholds to discount rates. A purchaser with discount rate r accepts capital C against annual savings S over measure life L when $S \cdot \mathrm{AF}(r,L) \ge C$ with $\mathrm{AF}(r,L) = (1-(1+r)^{-L})/r$. Since simple payback is $T = C/S$, the acceptance condition is

$$T \;\le\; \mathrm{AF}(r,L) \tag{27}$$

so the annuity factor is itself the maximum acceptable payback, and a bin value set to $\mathrm{AF}(r, L)$ represents a purchaser segment with discount rate r. The approximation $T_{\max} \approx 1/r$ is the $L \to \infty$ limit of (27) and overstates the threshold at finite equipment lives. With bin values set by (27), the capital charge C/p is the annualized cost of capital at that segment's discount rate, because $1/\mathrm{AF}(r, L)$ is the capital recovery factor at rate r.

Efficiency and fuel-switching measures carry no payback-acceptance allocation. They are costed per unit of annual energy savings (Section 8.8).

8.7 Technology linkage and ratio constraints

Technologies adopted jointly, such as the heating and cooling functions of one heat pump, which are sold in separate subsectors, are coupled by requiring the multiplier-weighted sales of all source technologies mapping to a target to equal the target's weighted sales:

$$\sum_{(g',\tau') \to (g,\tau)} \mathrm{Sales}_{g',\tau'}\, m_{g'}\, \lambda_{g',\tau'} \;=\; \mathrm{Sales}_{g,\tau}\, m_{g} \tag{28}$$

Where a source technology's linked target is not optimized, the source is fixed at its EP baseline.

Policy shares and technology mandates are imposed by a ratio constraint balancing weighted service-demand deviations (24) over two technology subsets:

$$\sum_{\tau \in \mathcal{A}} \mathrm{SD}_{y,g,\tau}\, v_{c,g,\tau,\mathcal{A}} \;+\; \sum_{\tau \in \mathcal{B}} \mathrm{SD}_{y,g,\tau}\, v_{c,g,\tau,\mathcal{B}} \;=\; 0 \tag{29}$$

Sales-share standards, minimum technology shares, and prescribed transitions are specified through the coefficient sets.

8.8 Efficiency and fuel-switching measures

Efficiency measures reduce consumption of a carrier; fuel-switching measures reduce consumption of one carrier and add consumption of another. The exogenous demand of Section 8.1 reflects full adoption of the measures in the EP case, and the measure variable $M_{m,y,z}$ of (23) is the fraction not adopted. The not-adopted savings of a measure may not fall from one optimization year to the next:

$$\sum_{\beta} M_{m,y,z}\, \varsigma_{y,z,m,\beta} \;+\; \psi_{m,y,z} \;\geq\; \sum_{\beta} M_{m,\,y-1,\,z}\, \varsigma_{y-1,z,m,\beta} \tag{30}$$

where ψ is a penalized slack. The condition applies to savings, not to the measure fraction. The first year is exempt. The slack accommodates a measure whose per-unit savings declines as its baseline becomes more efficient, in which case the prior year's absolute savings cannot be sustained.

Measure cost is accounted by vintage. The not-adopted savings added in year v, equal to the increase in $\sum_{\beta} M\varsigma$ from the preceding year, form a tranche that is credited in every year from v onward at the measure's cost per unit of annual savings for vintage v. The credit enters the objective with a negative sign: not adopting a measure avoids its cost and restores the demand it would have saved. Condition (30) keeps each tranche non-negative except through the slack.

8.9 Supply-side coupling and cost accounting

Technology decisions enter the supply-side balances through technology efficiency and profile. For electric technologies, the service-demand deviation (24) is converted to an hourly electricity load deviation using a technology-specific hourly profile where defined and a subsector default otherwise:

$$L^{\mathrm{endo}}_{y,z,s,t} \;=\; \sum_{(w,j)} \big(1 + \ell_{z,w,y}\big) \sum_{(g,\tau)\,\in\,(z,w,j)} \mathrm{SD}_{y,g,\tau}\; m_g\, e_{g,\tau}\; u_{y,s,t,z,\tau,j} \tag{31}$$

where $e_{g,\tau}$ is the technology's electricity use per unit of service, u its profile on feeder j, and the inner sum runs over technologies of vintage at or before y. Technologies resolved at representative-day resolution spread their daily energy evenly over the hours of the day. Because SD is a deviation, L^{endo} can be negative, and the sum of exogenous and endogenous load is the load of the solved sales. For technologies consuming other carriers, the corresponding deviations $D^{\mathrm{endo},\beta}$ are computed at hourly, representative-day, or annual resolution according to each technology's balance resolution and enter (6) and the annual accounting identity. Endogenous load is assigned to distribution feeders through the technology profiles and enters the planning reserve (4) and the peak definition of Section 4.9.

In mode (b), demand-side costs enter the objective of Section 2.3: technology capital cost, fixed operating cost, and incentives; efficiency and fuel-switching measure cost; electricity delivery cost; and load-shifting cost where load shifting is used. Penalty terms price the envelope and non-decreasing-savings slacks. Societal-cost variants of technology capital cost and incentives are maintained alongside private-cost variants, permitting reporting on either basis. For an optimized subsector, the levelized demand-side cost reported from EP is limited to stock built before the horizon, and the cost of sales in the optimization years is computed from the solved sales.

8.10 Mode-specific variables and constraints

The following variables, constraints, and objective terms exist only in mode (b). In mode (a) the corresponding quantities are zero.

Element	Type	Reference
Sales by technology and vintage; sales by payback bin	Variable	(24), (26)
Service-demand deviation	Variable	(24)
Slack below the sales floor and above the sales ceiling	Variable	Section 8.5
Endogenous electric load, hourly and annual, by zone, sector, and feeder	Variable	(31)
Endogenous carrier demand at hourly, representative-day, and annual resolution	Variable	Section 8.9
Sales envelope floor and ceiling	Constraint	Section 8.4
Service-demand definition and conservation	Constraint	(24), (25)
Payback-bin allocation	Constraint	(26)
Technology linkage, and fixing of sources with non-optimized targets	Constraint	(28)
Demand technology ratio constraints	Constraint	(29)
Endogenous load and carrier demand definitions	Constraint	(31)
Measure non-decreasing condition and vintage cost accounting	Constraint	(30), Section 8.8
Demand technology capital, fixed operating, and incentive cost; measure cost; envelope and measure-slack penalties	Objective term	Section 2.3

The measure fraction M and its slack are defined as variables whenever the measure switch is on, and the measure terms (23) then enter the balances; the measure constraints and cost terms in the table are built with the demand-optimization constraint set.

References

Williams, J. H., Jones, R. A., Haley, B., Kwok, G., Hargreaves, J., Farbes, J., & Torn, M. S. (2021). Carbon-neutral pathways for the United States. *AGU Advances*, 2(1), e2020AV000284. https://doi.org/10.1029/2020AV000284