

Vision and Progress towards a
Foundation Models for
the electric grid (GridFM)

Hendrik F. Hamann
Thomas Brunschweiler

Our GridFM community has been growing

✓ “IBM Yorktown” workshop



✓ “Imperial College” workshop



✓ GridFM
Linux Foundation
Energy project



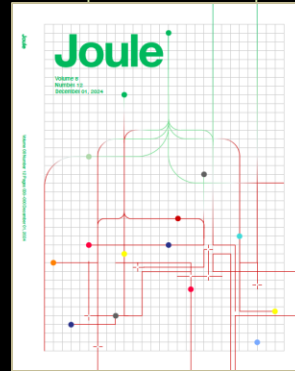
Mar 24

✓ Founding of
GridFM working group



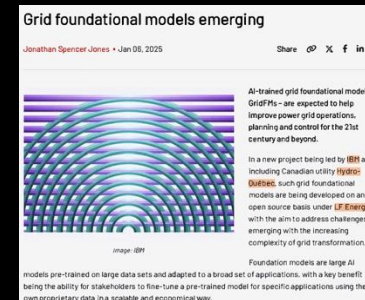
Jun 24

✓ Joule GridFM
Perspective Paper



Dec 24

✓ Communication
outreach



Mar 25

GridFM Conference
@Argonne



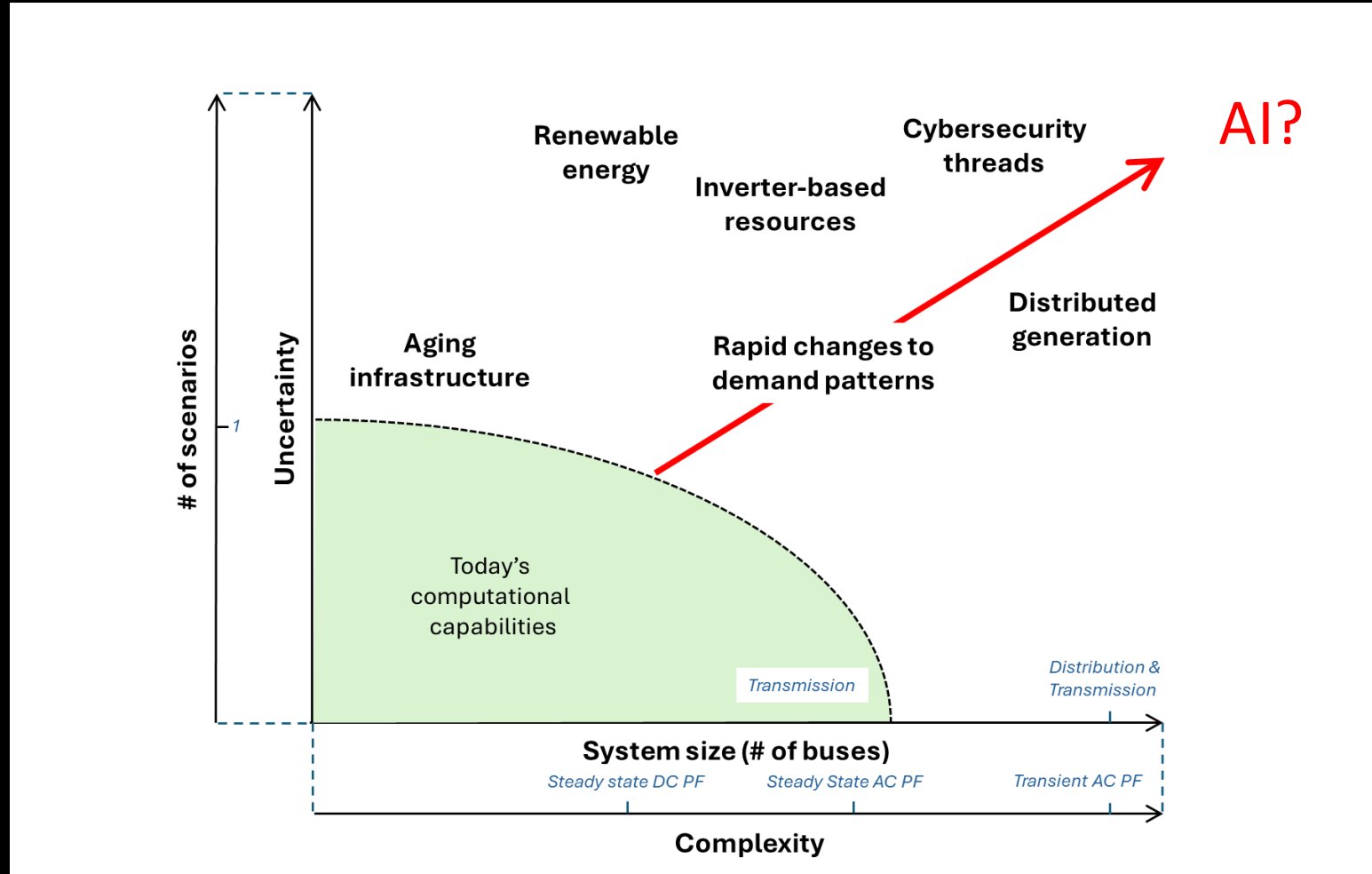
First version
of GridFM ?

Subgroups launched ✓

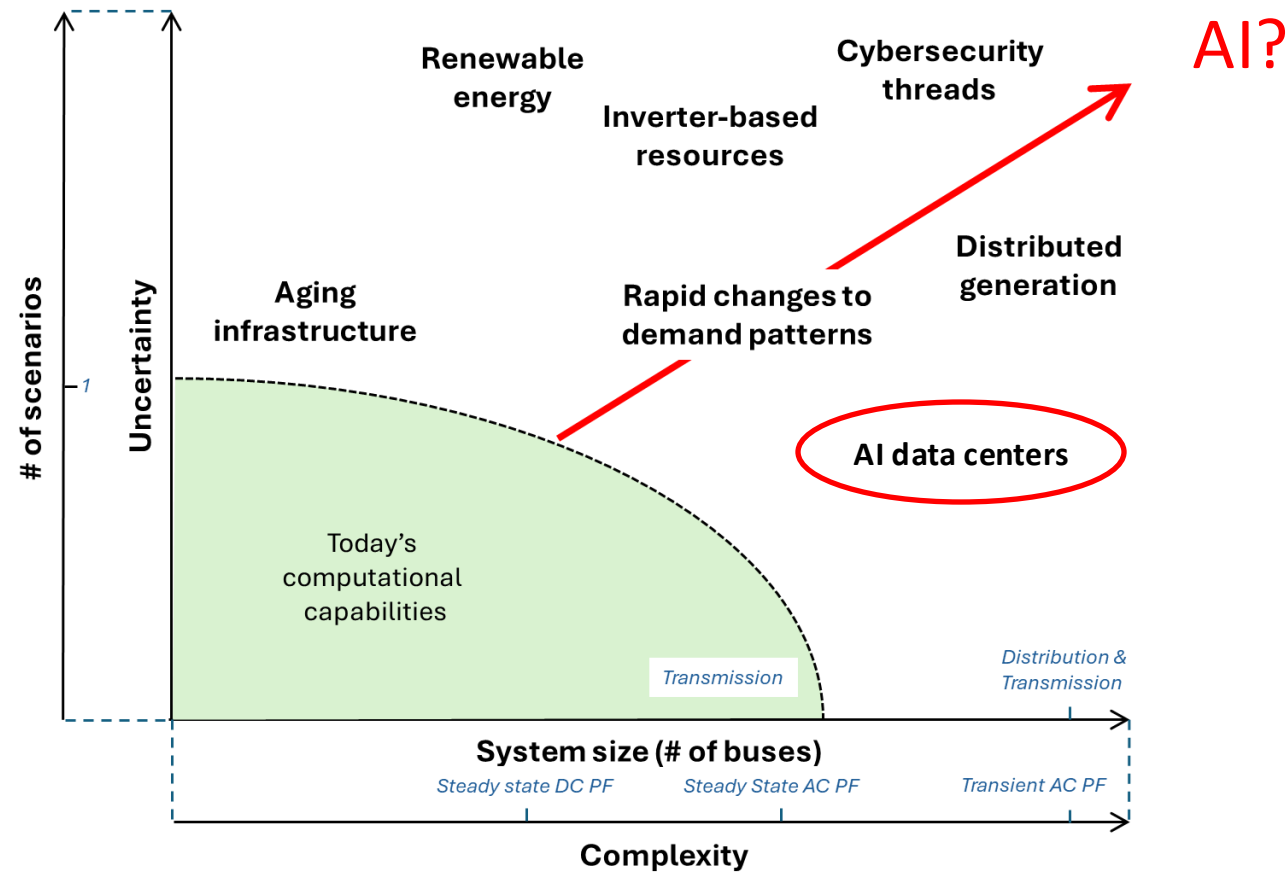
Technology, Governance, Collaboration

Why all this interest?

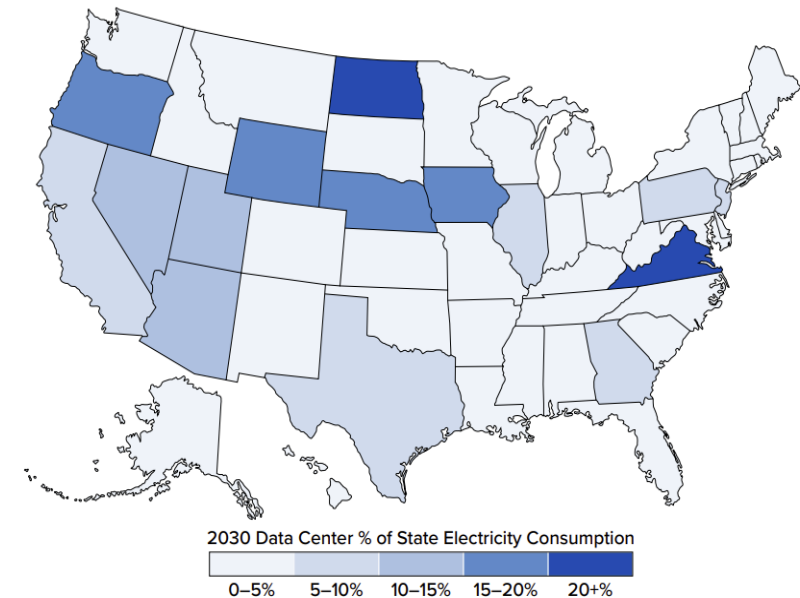
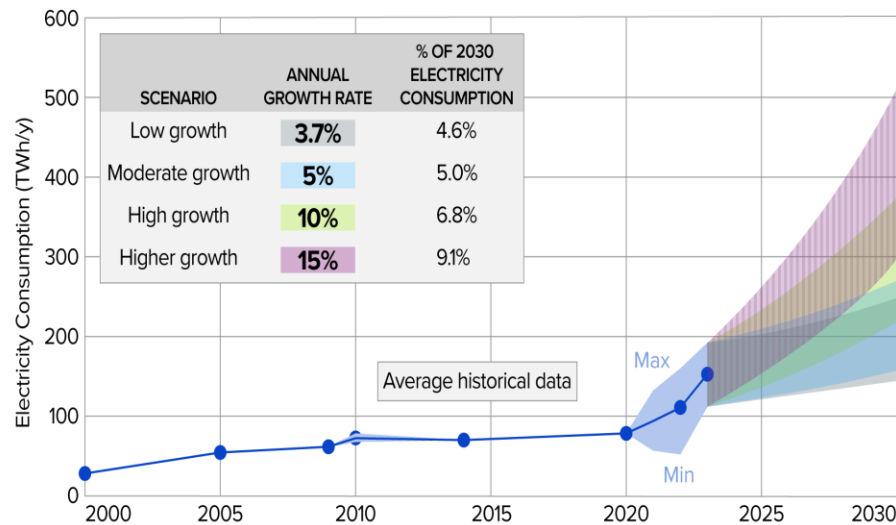
GridFM is an opportunity to cope with increasing complexity and uncertainty of the power grid



But wait a minute...
AI itself poses yet another major
challenges to the power grid



AI Data Center Power Demands are growing rapidly adding to the complexity and uncertainty of the electric grid



- ✓ AI pretraining as a dispatchable load (distributed energy resource)
- ✓ Inference as a non-dispatchable (consumer driven) load

Why did we think that AI Foundation Models are a good idea for power grids?

AI Foundation Models excel in

- Predicting the next data
"token" based on context

-
- Homogenization
and adaptability

-
- Massive Acceleration
of simulations

Why did we think that AI Foundation Models are a good idea for power grids?

AI Foundation Models excel in

- Predicting the next data "token" based on context

-
- Homogenization and adaptability

-
- Massive Acceleration of simulations

Enabling tech

Attention / transformer architectures

Self-supervision, big data & pervasive compute

Deep networks to map directly inputs to outputs

Why did we think that AI Foundation Models are a good idea for power grids?

AI Foundation Models excel in

- Predicting the next data "token" based on context

-
- Homogenization and adaptability

-
- Massive Acceleration of simulations

Enabling tech

Attention / transformer architectures

Self-Supervision, big data & pervasive compute

Deep networks to map directly inputs to outputs

Previous Work (Geospatial FMs)

- Excellent spatial & temporal reconstruction performances

-
- Adaptable to multiple downstream applications
 - Improved data efficiency and accuracy

-
- 10^5 acceleration for weather forecasting compared to simulations

Why did we think that AI Foundation Models are a good idea for power grids?

AI Foundation Models excel in

- Predicting the next data "token" based on context

-
- Homogenization and adaptability

-
- Massive Acceleration of simulations

Enabling tech

Attention / transformer architectures

Self-Supervision, big data & pervasive compute

Deep networks to map directly inputs to outputs

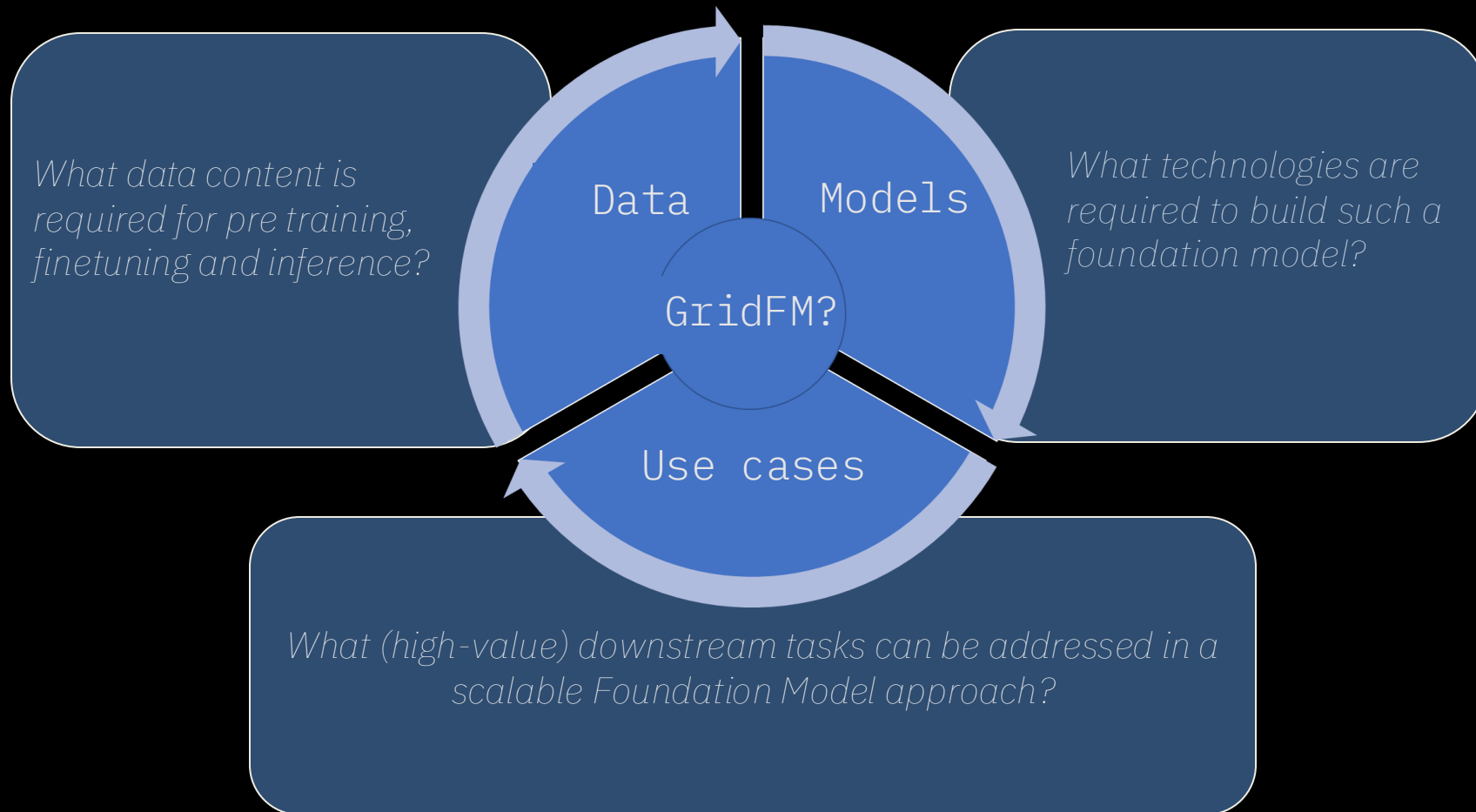
Opportunities to power grids

- Power flow
- Load and renewable forecasts
 - State estimation
 - ...

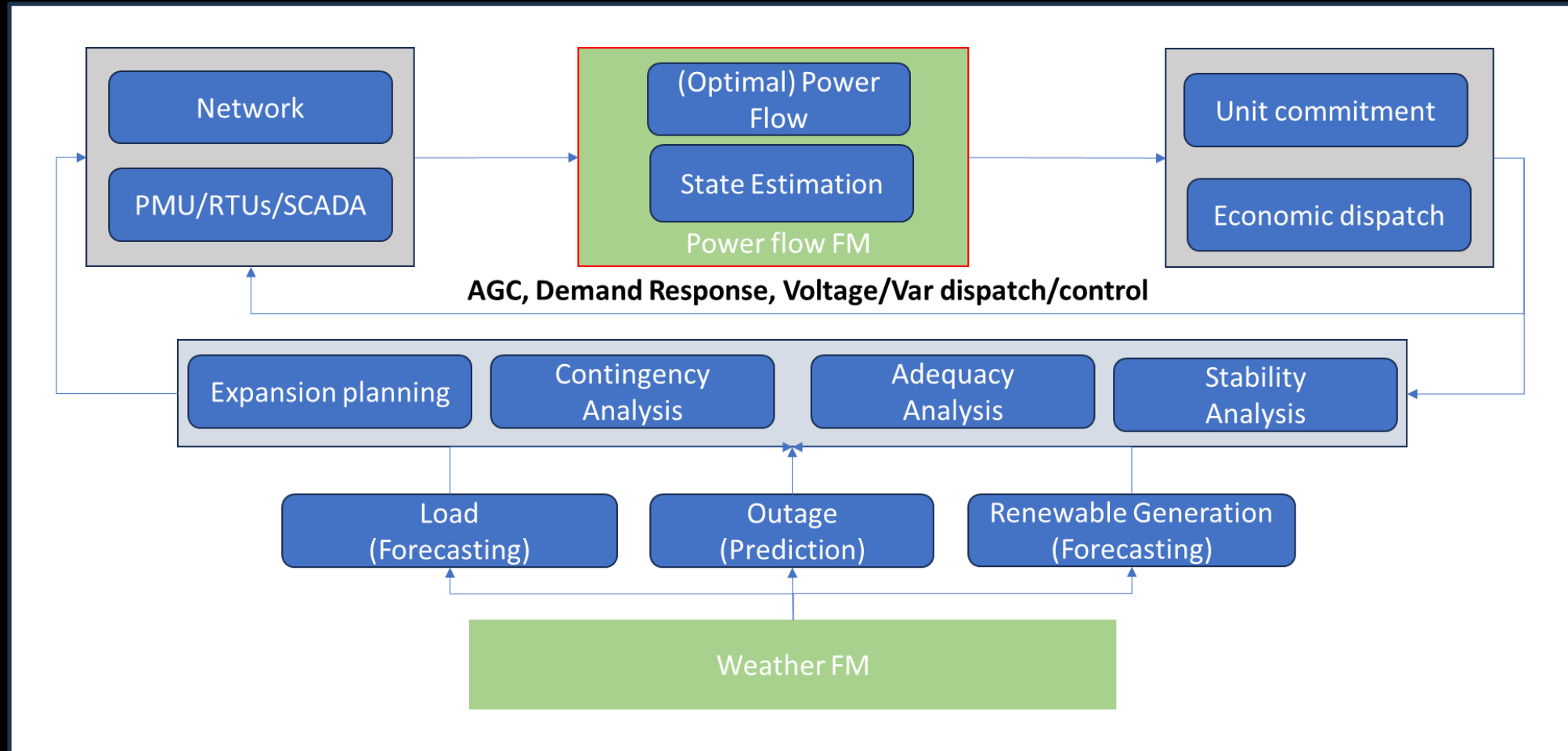
-
- Scaling across different grids
 - Collaboration platform
 - ...

-
- (N-k) contingency analysis with $k > 1$
 - Co-simulations for transmission & distribution
 - Transients and Optimal Power flow
 - Accelerated interconnection studies
 - ...

Steps to define GridFM-v0



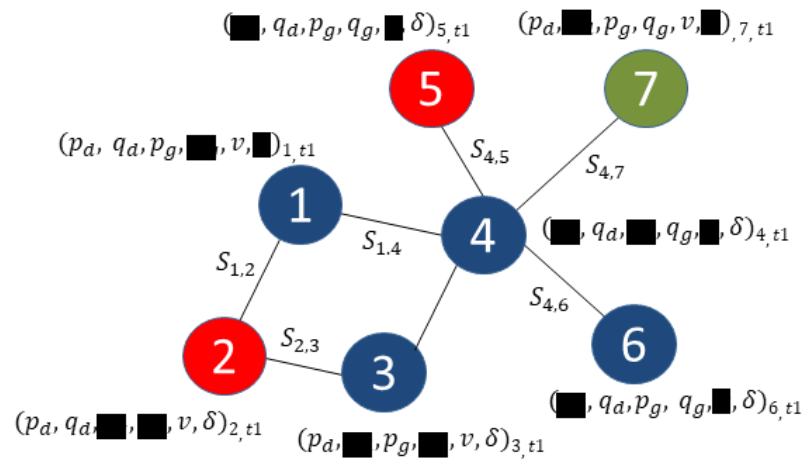
The answer we came up with: FM pretrained for power flow



Power flow estimations are at the core of operating, controlling and planning the electric grid

Humbled beginnings: GridFM-v0 pretrained on power flow data

Masked Graph \mathcal{G}'

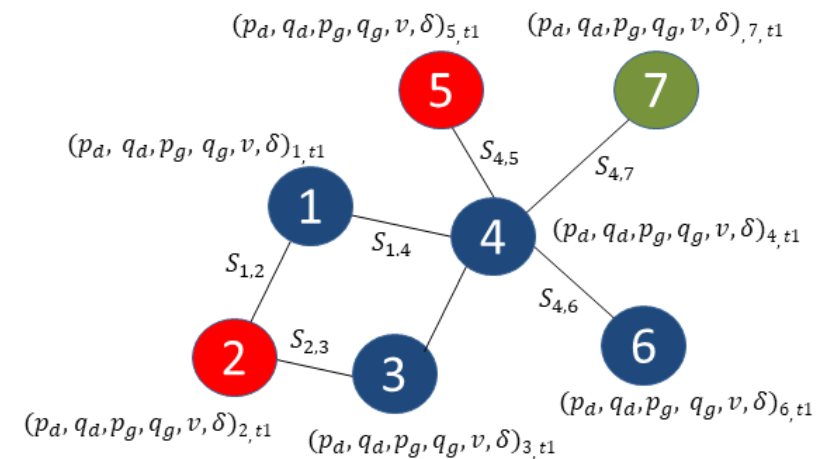


- Load bus
- Generator bus
- Slack bus
- Line

Pre-training
 $\mathcal{G} = f(\mathcal{G}')$

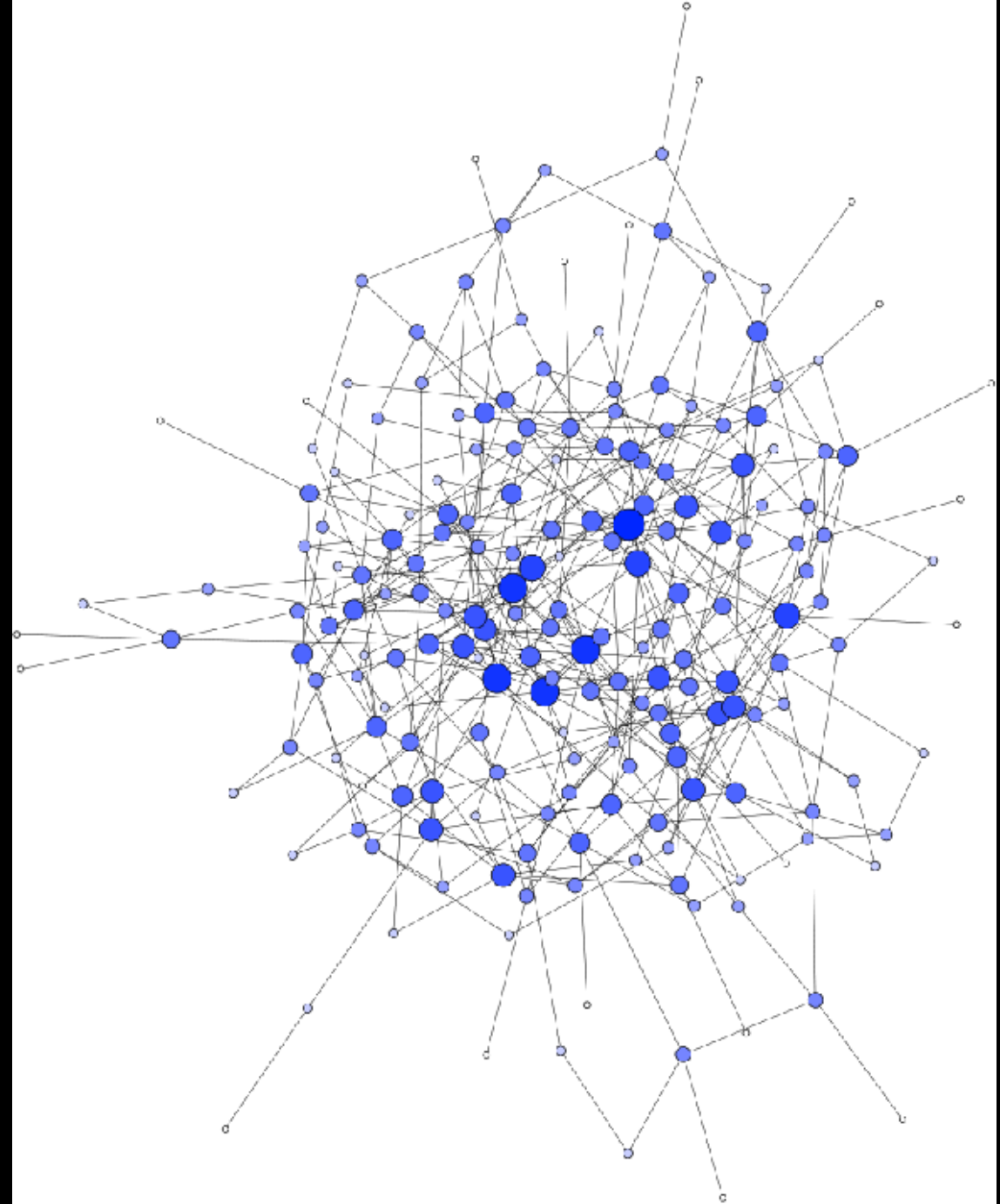
$m_a(\mathcal{G}) = \mathcal{G}'$
Masking

Reconstructed Graph \mathcal{G}



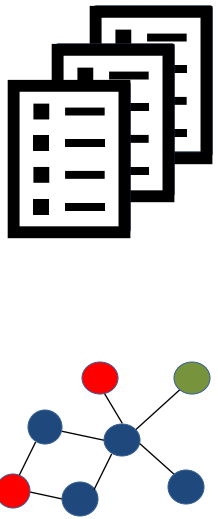
We are starting with public grids...

- IEEE 14-Bus System
- IEEE 24-Bus System
- IEEE 30-Bus System
- IEEE 39-Bus System
- IEEE 57-Bus System
- IEEE 118-Bus System
- IEEE 300-Bus System
- GB network
- ACTIVSg200
- ACTIVSg500
- ACTIVSg2000
- ACTIVSg10k
- ACTIVSg25k
- ACTIVSg70k
-

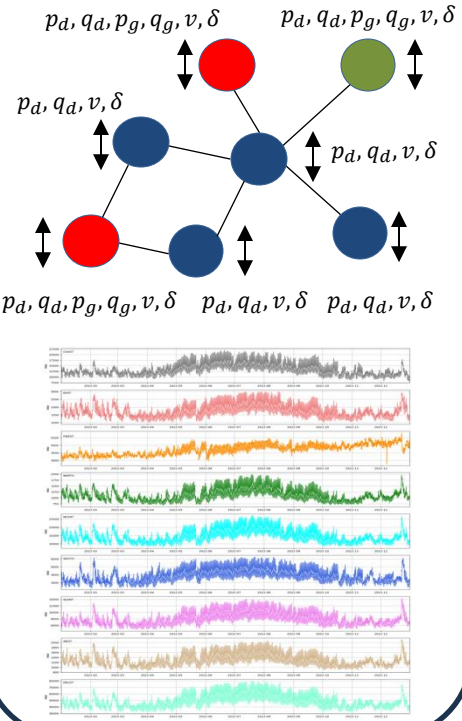


Pretraining Data Generation for GridFM-v0 from public grids

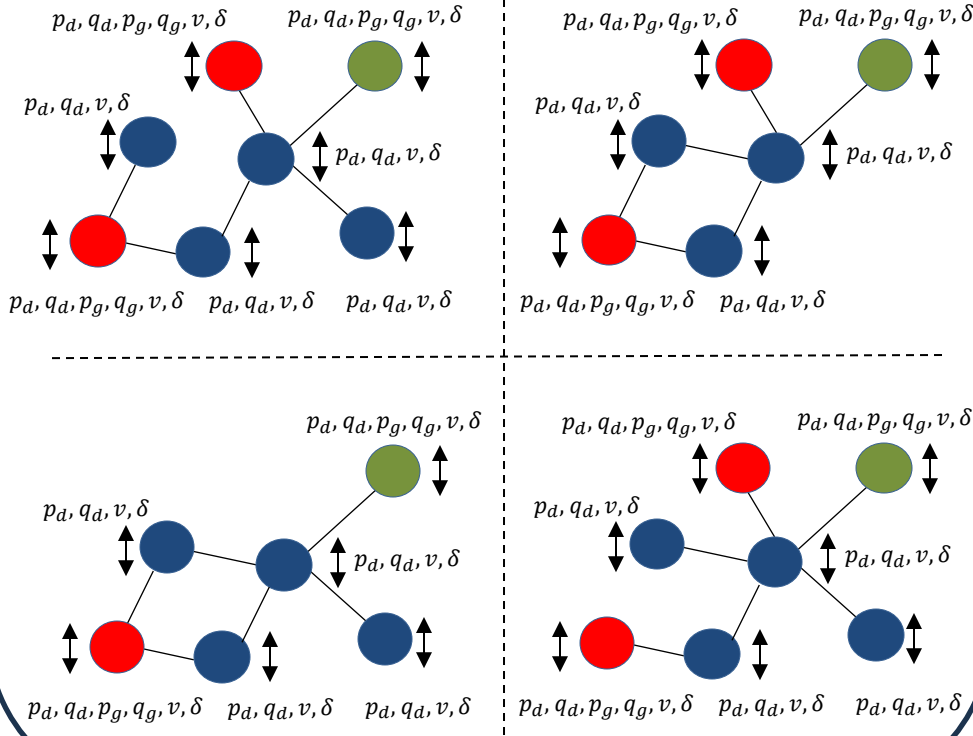
Load & Case files



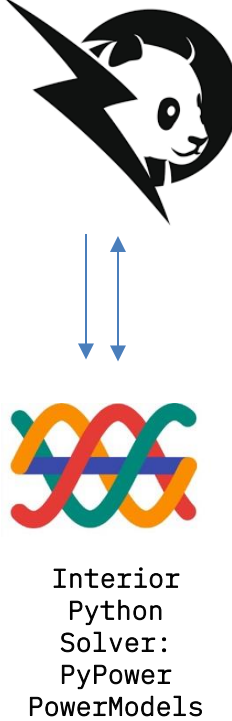
Grid-level, bus-level and load perturbations



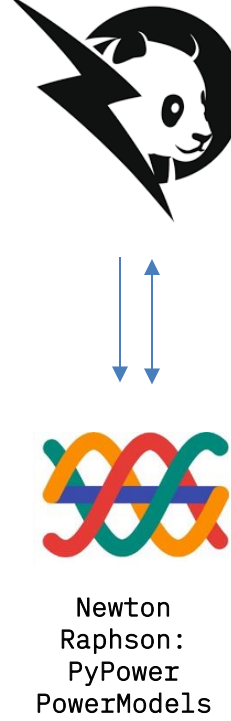
Grid topology perturbation



Solve OPF & generation dispatching



Solve Power Flow



Self-supervised pretraining with Masked Autoencoder

Data

Data normalization

Base MVA. p.u. system

Masking

Each node feature is being masked with 0.5 probability

GridFM-v0 pretraining

TransformerConv (Graph Neural Network)

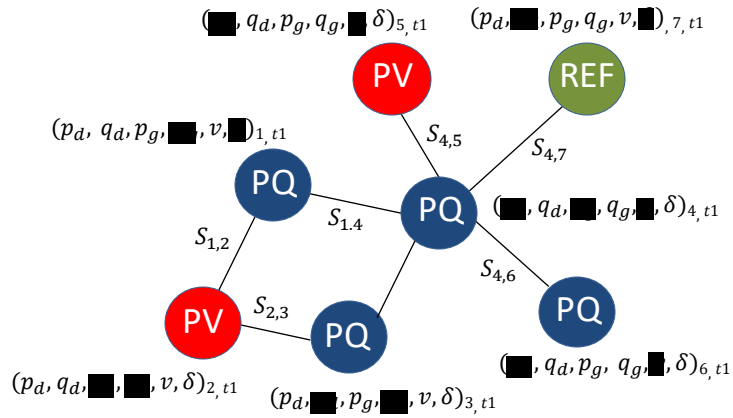
Physics-informed loss functions

$$S_i = S_i^{gen} - S_i^{load} = S_i^{trans}$$

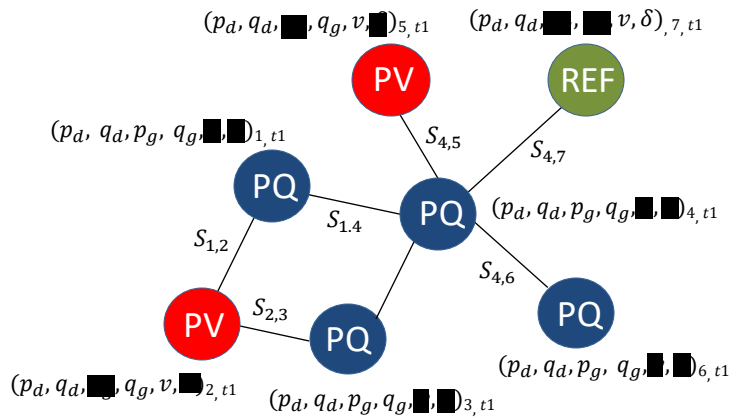
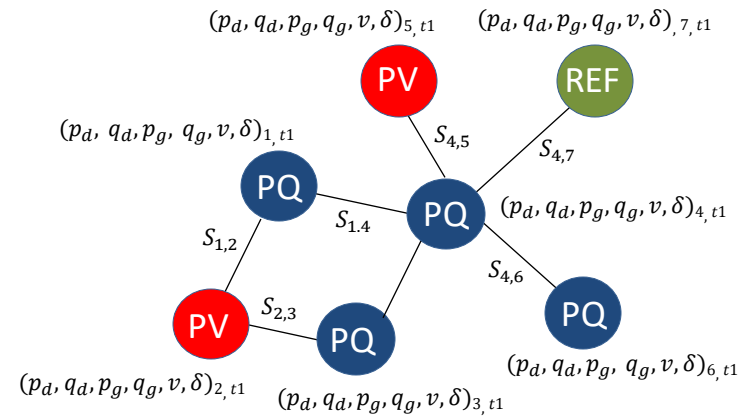
$$L = w_1 L_{MSE} + w_2 L_{PBE}$$

Linear combination of MSE and power loss

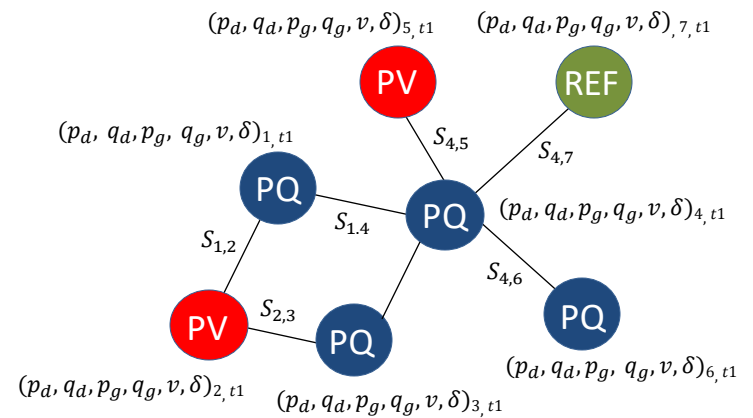
Masking for pretraining and finetuning



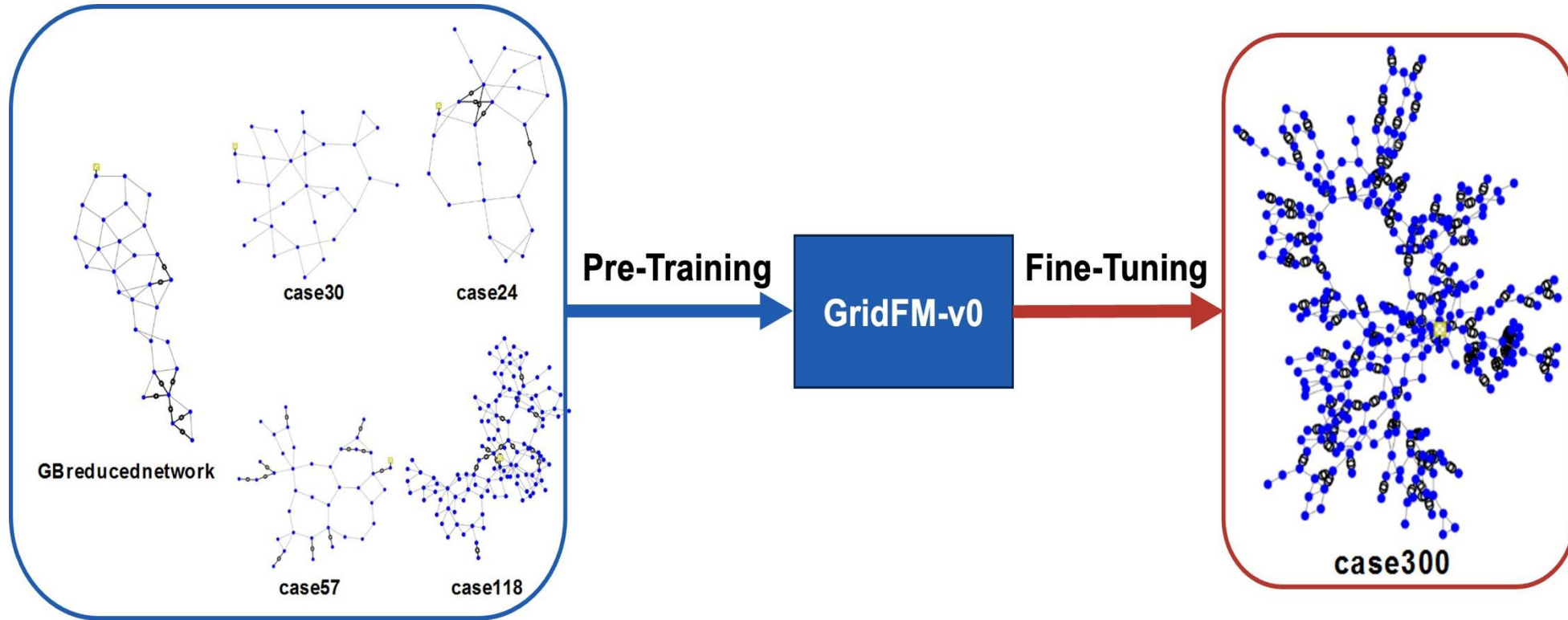
Pretraining



Finetuning

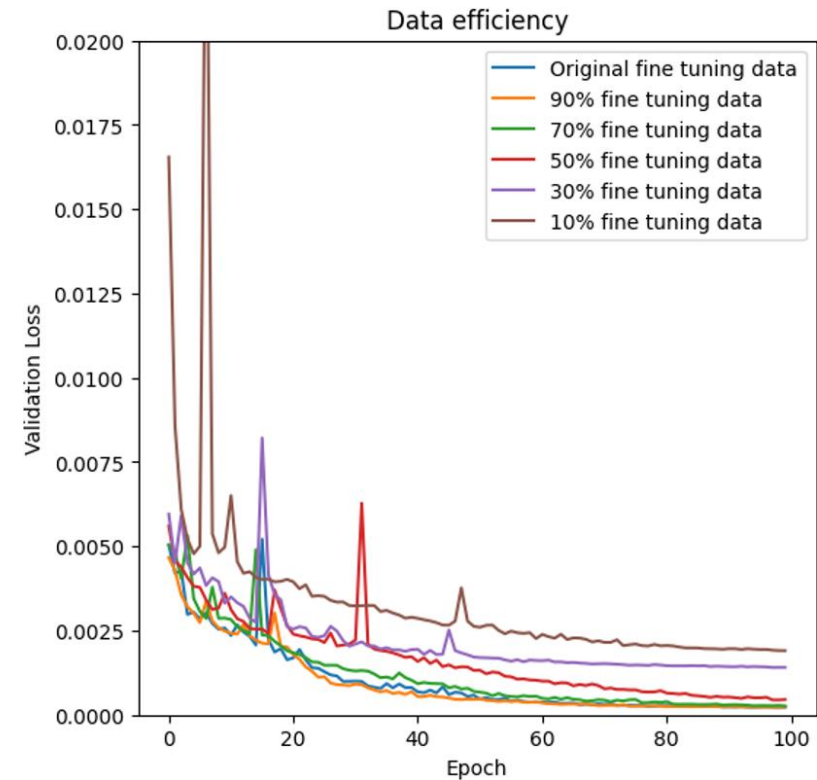
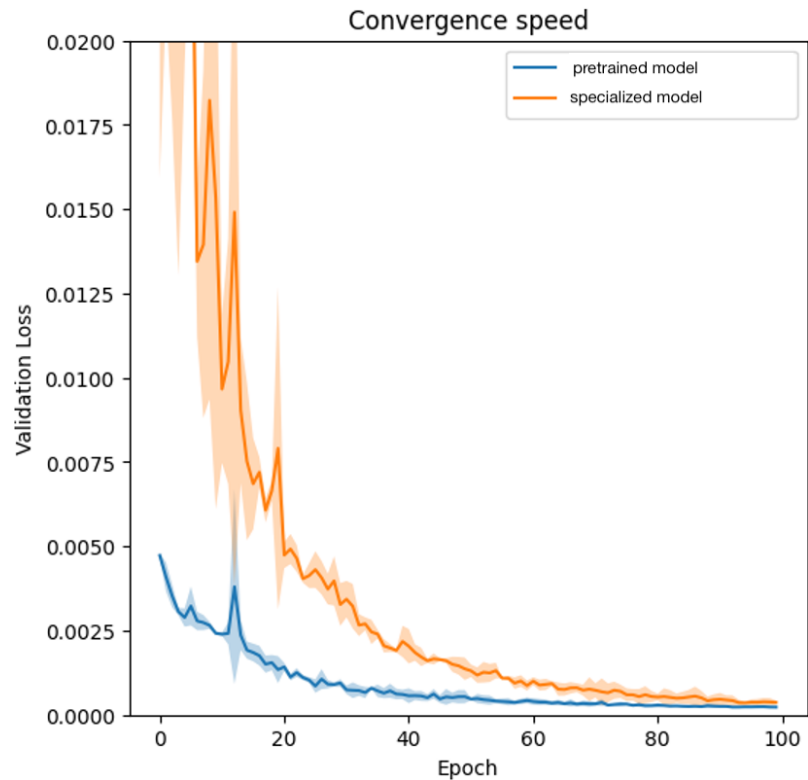


Can GridFM-v0 generalize to new grids?



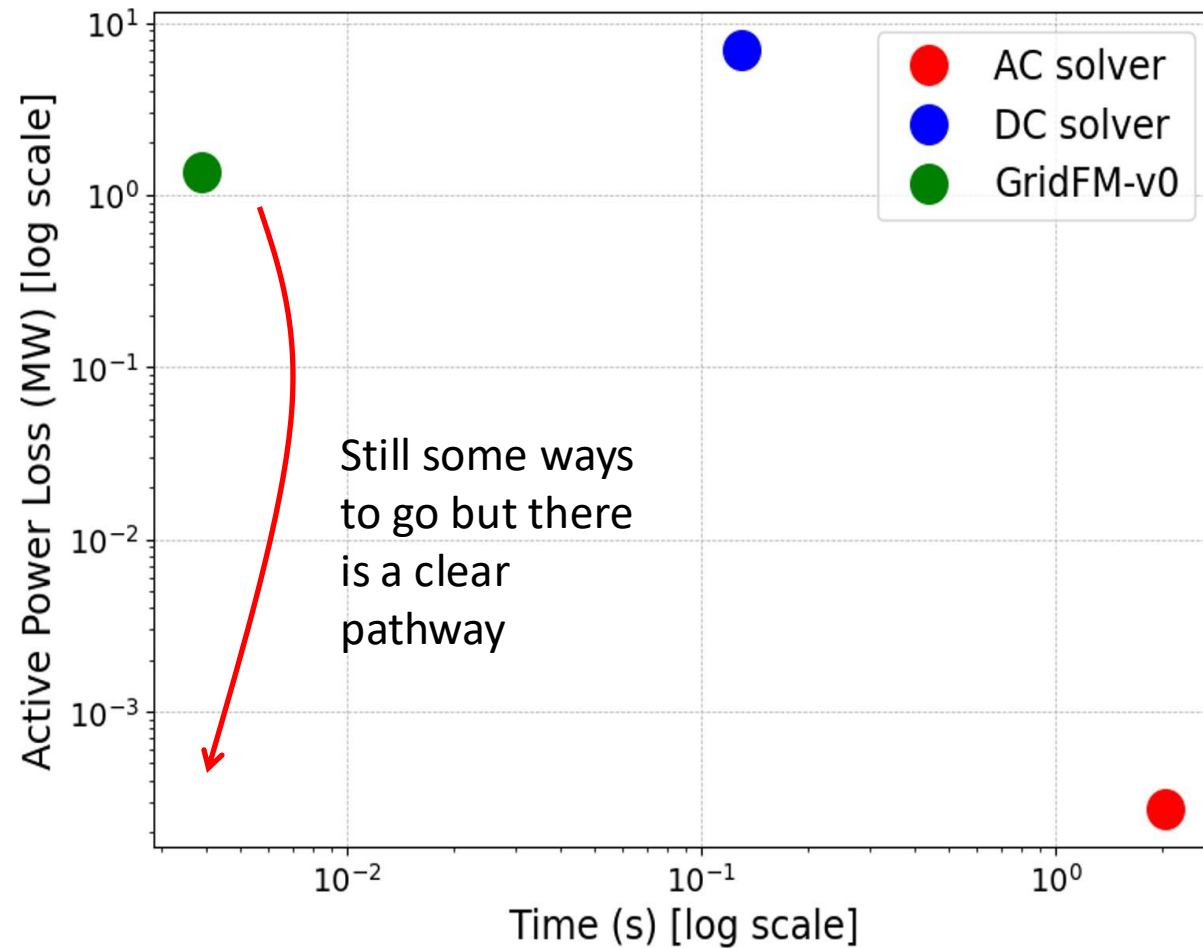
- ✓ Pre-training on five small networks (total of 75k load scenarios)
- ✓ Fine-tuning on larger network (18k load scenarios)

Accelerated convergence and improved data efficiency achieved by pretraining

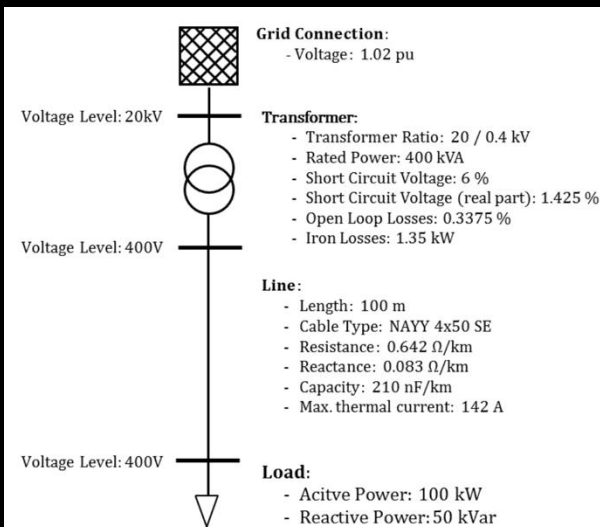


- GridFM-v0 converges faster during finetuning than specialized supervised model
- Readily adaptable to 2000+ grids in the US alone

Early results - GridFM emulates power flow faster
...and even better than DC solvers



Can't I do all this with ChatGPT?



ChatGPT o1

We want to learn more about a three-bus electrical network. The first bus has a voltage level of 20kV and is connected to another grid with 1.02 pu voltage. The second and third bus have a voltage level of 0.4kV. There is a load connected to bus 3 with 100kW of active power and 50kVar of reactive power. There is a transformer, which connects between bus 1 and bus 2, which has a ratio of 20 / 0.4 kV. The maximum rated power is 400 kVA. The short circuit voltage is 6% with the real part of 1.425%. The open loop losses are 0.3375% and the iron losses are 1.35 kW.

There is a cable line of 100m between bus 2 and 3. The cable type is NAVY 4x50 SE. The resistance, reactance, capacity of the cable is 0.0642 Ohm, 0.0083 Ohm and 21 nF, respectively. The maximum thermal current is 142 A. Can you compute the voltage magnitude and angle at node 3?

Reasoned about voltage at bus 3 for 4m 57s

Below is a straightforward way to arrive at an approximate "one-shot" load-flow solution for the voltage at the 0.4 kV load bus (bus 3). The key idea is to refer everything to the low-voltage (LV) side of the transformer, treat the external 20 kV system (bus 1) as an ideal 408 V source* in series with the transformer impedance, then include the cable impedance out to bus 3. You can then apply standard radial "back-of-the-envelope" formulas for voltage-drop with a lagging-power-factor load.

1) System data referred to the 0.4 kV side

1. Nominal LV voltage:

$$V_{\text{nom, LV}} = 400 \text{ V (line-line)}.$$

Because the 20 kV side is at 1.02 p.u., i.e. 20.4 kV, the no-load LV side would be at

$$20.4 \text{ kV} / 50 = 408 \text{ V (line-line)}.$$

So, for load-flow purposes, we can model the upstream (HV) system as a stiff 408 V source on the LV side, if there is no transformer voltage drop. Next we add in the transformer series impedance.

2. Transformer short-circuit impedance (referred to LV side):

- Rated power: $S_T = 400 \text{ kVA}$.
 - Percent impedance: $Z_{\%} = 6\%$, with the real part $R_{\%} = 1.425\%$.
 - On the LV side base, the absolute transformer impedance is
- $$Z_T = 0.06 \times Z_{\text{base, LV}}, \quad Z_{\text{base, LV}} = \frac{(0.4 \text{ kV})^2}{0.4 \text{ MVA}} = 0.4 \Omega.$$

Hence

$$Z_T = 0.06 \times 0.4 = 0.024 \Omega$$

of which the real part is 0.01425 p.u. $\rightarrow 0.01425 \times 0.4 = 0.0057 \Omega$ and the reactive part is $\sqrt{(0.024)^2 - (0.0057)^2} \approx 0.0233 \Omega$.

3. Cable data (bus 2 to bus 3, 100 m):

- $R_c = 0.0642 \Omega$, $X_c = 0.0083 \Omega$.
- The capacitive current is negligible at this low voltage/short length, so we usually omit it in a simple radial load-flow.

Hence, on the LV side we have an "ideal" 408 V source in series with

$$Z_{\text{tx}} = 0.0057 + j 0.0233 \Omega \quad \text{and} \quad Z_{\text{cable}} = 0.0642 + j 0.0083 \Omega,$$

leading to bus 2, then the cable out to bus 3.

2) Load at bus 3

The 3-phase load is:

$$P = 100 \text{ kW}, \quad Q = 50 \text{ kvar} \quad \rightarrow \quad S = 100 + j 50 \text{ kVA} \approx 112 \text{ kVA}.$$

At about 400 V (line-line), the load current magnitude is

$$I_{\text{load}} = \frac{|S|}{\sqrt{3} V_{\text{LL}}} = \frac{112,000 \text{ VA}}{\sqrt{3} \times 400 \text{ V}} \approx 161 \text{ A}.$$

Because the load is inductive ($P/Q = 2$), the power factor is

$$\cos \varphi = \frac{P}{\sqrt{P^2 + Q^2}} \approx \frac{100}{112} \approx 0.894,$$

so $\varphi \approx 26^\circ$ (lagging).

3) Approximate radial drop calculation

A common approximate formula for the magnitude of voltage drop ΔV in a line or transformer with series impedance $R + jX$ is (for lagging power factor):

$$\Delta V \approx I (R \cos \varphi + X \sin \varphi).$$

Here:

- The total $R_{\text{tot}} = 0.0057 + 0.0642 = 0.0699 \Omega$.
- The total $X_{\text{tot}} = 0.0233 + 0.0083 = 0.0316 \Omega$.

Hence, with $I \approx 161 \text{ A}$ and $\cos \varphi = 0.894$, $\sin \varphi = 0.447$,

$$\Delta V_{\text{total}} = 161 (0.0699 \times 0.894 + 0.0316 \times 0.447).$$

Compute inside:

- $0.0699 \times 0.894 \approx 0.0625$,
- $0.0316 \times 0.447 \approx 0.0141$,
- sum ≈ 0.0766 ,
- multiply by 161 A $\approx 12.3 \text{ V}$.

Since the no-load LV side is 408 V, we subtract the drop of about 12-13 V. This gives a bus 3 voltage near:

$$V_3 \approx 408 \text{ V} - 12.3 \text{ V} \approx 395.7 \text{ V}.$$

In per-unit on a 400 V base, that is

$$\frac{395.7}{400} \approx 0.989.$$

That is only about a 1 percent dip from nominal 400 V.

The phase-angle shift across these short impedances (and at only ~1-2% voltage drop) is small—on the order of a few tenths of a degree of lag relative to the HV side. A more exact phasor iteration typically ends up with something like $\angle V_3 \approx -0.1^\circ$ or -0.2° referred to bus 1.

4) Representative final answer

Putting this all together:

- Bus 3 magnitude: about 395-396 V (line-line), i.e. 0.988-0.99 p.u. on a 400 V base.
- Bus 3 angle: about -0.1° to -0.2° relative to the HV-side reference (if we take bus 1 at 0°).

A more detailed, fully iterated AC load-flow would give very nearly the same result for such a small radial system.

* Why 408 V as the "ideal source"?

Since the HV side is at 20.4 kV (1.02 p.u. on a 20 kV base) and the nominal transformer ratio is 20 kV:0.4 kV = 50:1, no-load on the LV side would be at

$$\frac{20.4 \text{ kV}}{50} = 408 \text{ V}.$$

Hence one can model "bus 1" as a 408 V stiff source in series with the transformer short-circuit impedance (0.0057 + j0.0233 Ω on the LV side). Then the cable is in series to bus 3.

Unfortunately, wrong result but still impressive...

GridFM Roadmap: from POC to hardened implementation ... towards general grid intelligence (GGI)

Base-GridFM Extension

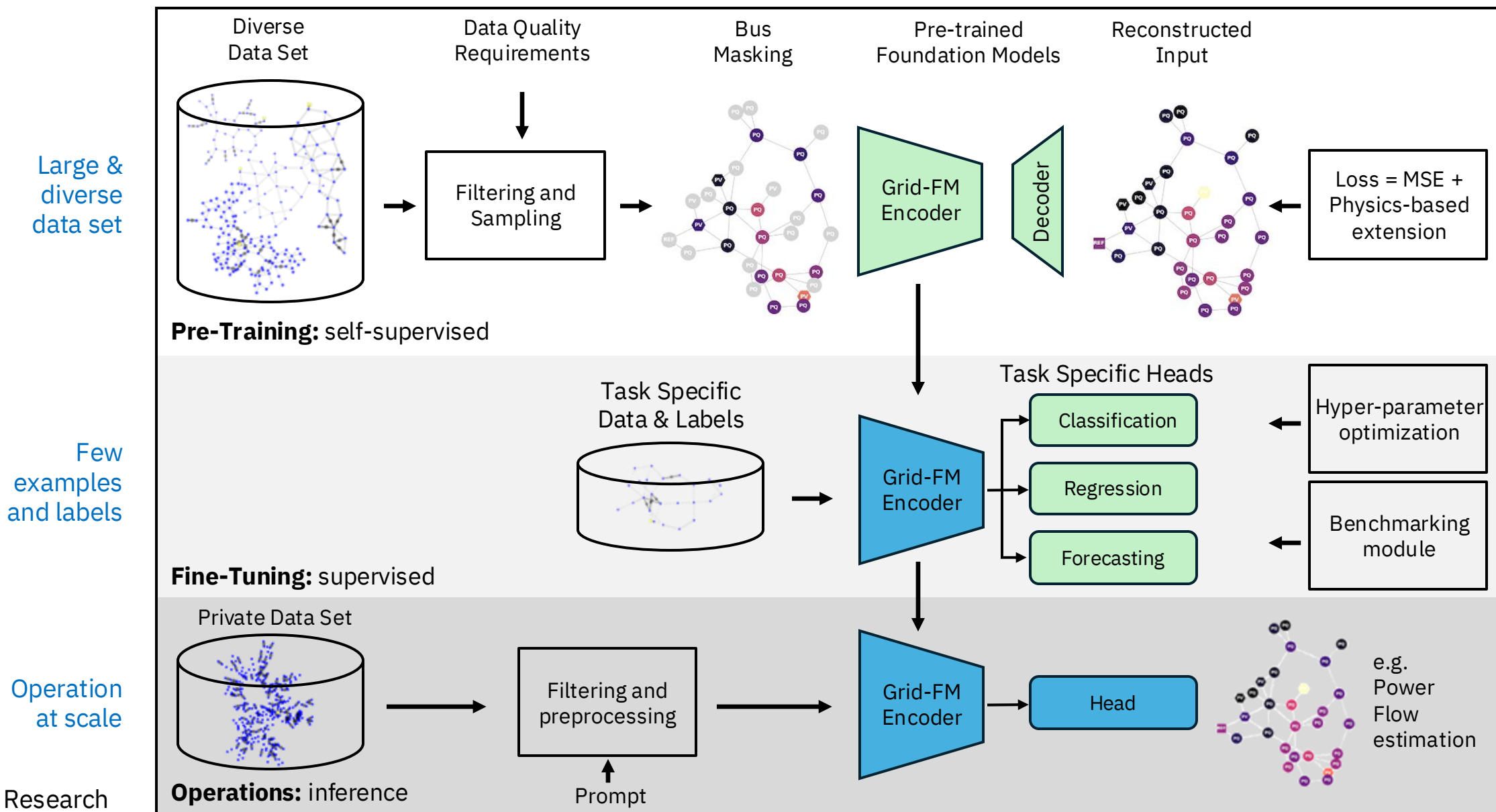
	Base-GridFM	Applications & Tasks	Real-World Data	Data & Model Hygiene	Tooling
Proof-of-concept	Reconstruction of bus-variables based on synthetic data with physics-informed loss for many topologies with local attention	Establish down-stream tasks and benchmarks: <ul style="list-style-type: none">• Power Flow• State Estimation• Contingency analysis• Cascading failure	Curate pre-training and fine-tuning data sets: <ul style="list-style-type: none">• diverse grids and operation scenarios incl. edge cases• sampling reshape training data distribution• labels for applications	Methods to share sensitive data: <ul style="list-style-type: none">• Generative models• Differential-privacy• Federated learning Defense against model attacks: <ul style="list-style-type: none">• Model inversion attack	Implement tools to support model up-take, dev. experience: <ul style="list-style-type: none">• low-code GridFM tooling• model life-cycle management• scalable inference
Proof-of-generalization					
Hardening					
Governance					
Operations					



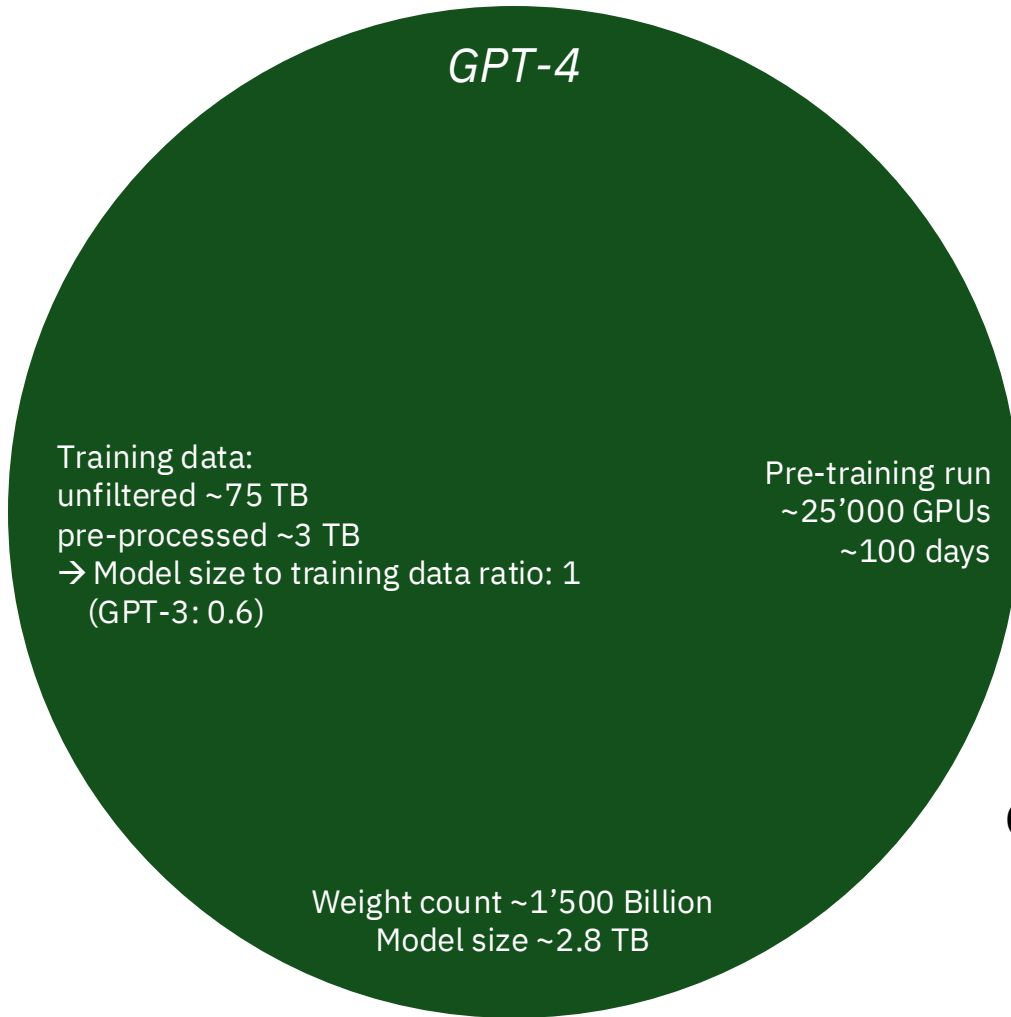
Demonstrated GridFM KPI's:

- Reconstruction capability
- Generalization across topologies
- Inference speed-up
- Data efficient fine-tuning

Grid Foundation Model Life-Cycle



LLM vs. Scientific FMs: same-same but different



Large Language Model

Attributes of Scientific FMs :

- i) They are orders of magnitude small compared to LLMs,
- ii) they do not need 10's of thousands of GPUs to be trained and
- iii) their training data volume is much larger than the model size – they may not have the capacity to memorize “all” the data.
- iv) The fine-tuning of a Scientific FM requires typically only one GPU!

600 Million ● *SAM*

Vision FM

30 Billion ● *MovieGen*

Gen-Video FM

50 Million ● *GridFM*
1GPU
0.001 model/train-data

Electric Grid FM

700 Million ● *Chronos*

Time-Series FM

2.3 Billion ● *Prithvi-WxC*
64 GPUs
0.001 model/data

Weather FM

Privacy & Data Protection Methodologies

Anonymization of Sensitive Data

Differential Privacy

Perturb sensitive data while maintaining statistical utility

Database

Name	Disease
Bob	y
Alice	n
Otto	n

Was Alice tested positive?



Heads:
true answer

↓
no

Tail:
Depends on
second coin flip

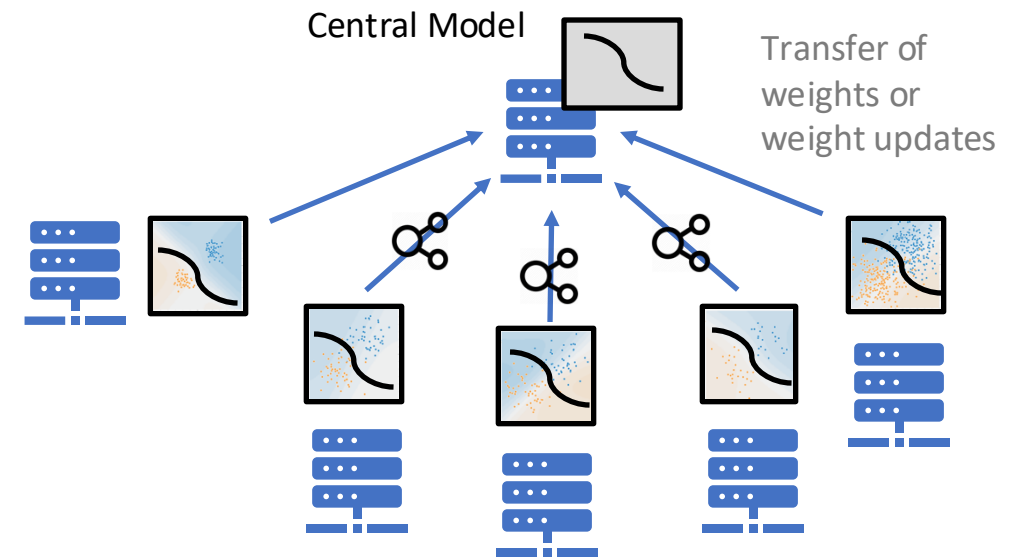
↓
Heads:
yes

↓
Tail:
no

Mitigate Transfer of Sensitive Data

Federated Learning

Distribute Model Training



GridFM Roadmap: from POC to hardened implementation ... towards general grid intelligence (GGI)

Base-GridFM Validation

	Base-GridFM	Applications & Tasks	Real-World Data	Data & Model Hygiene	Tooling
Proof-of-concept	Reconstruction of bus-variables based on synthetic data with physics-informed loss for many topologies with local attention: + topology agnostic + speed-up + performance increase to DC + data efficiency	Establish down-stream tasks and benchmarks: <ul style="list-style-type: none">• Power Flow• State Estimation• Contingency analysis• Cascading failure	Curate pre-training and fine-tuning data sets: <ul style="list-style-type: none">• diverse grids and operation scenarios incl. edge cases• sampling reshape training data distribution• labels for applications	Methods to share sensitive data: <ul style="list-style-type: none">• Generative models• Differential-privacy• Federated learning Defense against model attacks: <ul style="list-style-type: none">• Model inversion attack	Implement tools to support model up-take, dev. experience: <ul style="list-style-type: none">• low-code GridFM tooling• model life-cycle management• scalable inference
Proof-of-generalization					
Hardening					
Governance					
Operations					

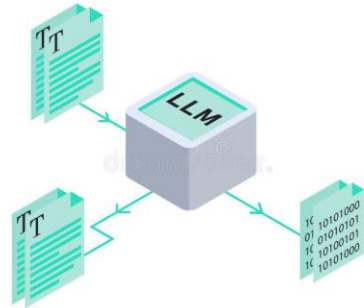
GridFM Family Extension

Base-GridFM+	Forecast-GridFM
Advanced model architectures: <ul style="list-style-type: none">• Alternative data representation & tokenization• Scalable local-global attention	Temporal reconstruction: <ul style="list-style-type: none">• Look-ahead power flow & state Model coupling & multi-modal data (time-seriesFM, weatherFM, LLM): <ul style="list-style-type: none">• Load & renewable generation forecasting

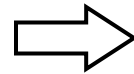
Foundation Model Coupling & Workflows

Large-Language Model

Feature mining from news & social media

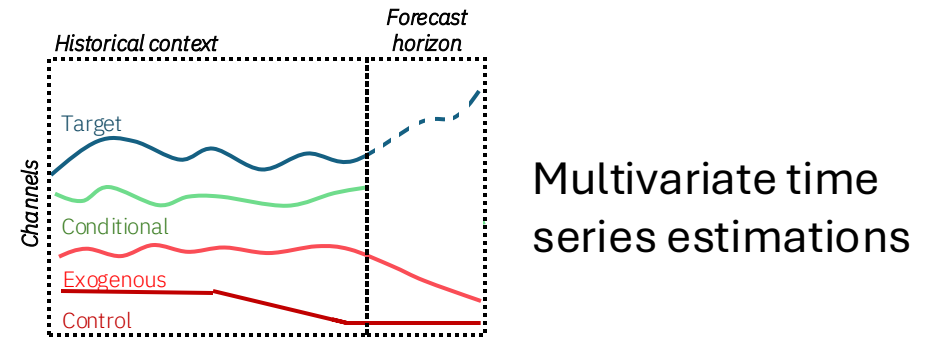


Advanced text understanding



Exogenous factors

Time-Series Foundation Model



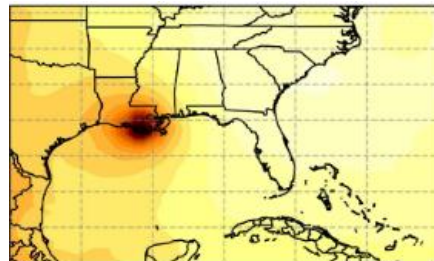
Multivariate time series estimations

Zero-shot capability



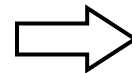
Weather Foundation Model

- Now-casting
- Down-scaling
- Extreme events

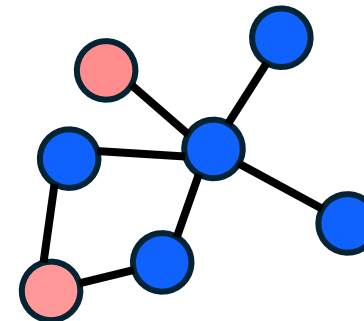


10'000x speed-up

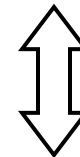
Renewable forecasting



Grid Foundation Model



Grid related forecasting tasks



Load forecasting

GridFM Roadmap: from POC to hardened implementation ... towards general grid intelligence (GGI)

Base-GridFM Validation

	Base-GridFM	Applications & Tasks	Real-World Data	Data & Model Hygiene	Tooling
Proof-of-concept	<p>Reconstruction of bus-variables based on synthetic data with physics-informed loss for many topologies with local attention:</p> <ul style="list-style-type: none"> + topology agnostic + speed-up + performance increase to DC + data efficiency 	<p>Establish down-stream tasks and benchmarks:</p> <ul style="list-style-type: none"> • Power Flow • State Estimation • Contingency analysis • Cascading failure 	<p>Curate pre-training and fine-tuning data sets:</p> <ul style="list-style-type: none"> • diverse grids and operation scenarios incl. edge cases • sampling reshape training data distribution • labels for applications 	<p>Methods to share sensitive data:</p> <ul style="list-style-type: none"> • Generative models • Differential-privacy • Federated learning <p>Defense against model attacks:</p> <ul style="list-style-type: none"> • Model inversion attack 	<p>Implement tools to support model up-take, dev. experience:</p> <ul style="list-style-type: none"> • low-code GridFM tooling • model life-cycle management • scalable inference
Proof-of-generalization			Hardening	Governance	Operations

GridFM Family Extension

Base-GridFM+	Forecast-GridFM	Optimal-GridFM	HighRes-GridFM	Advisor-GridFM
<p>Advanced model architectures:</p> <ul style="list-style-type: none"> • Alternative data representation & tokenization • Scalable local-global attention 	<p>Temporal reconstruction:</p> <ul style="list-style-type: none"> • Look-ahead power flow & state <p>Model coupling & multi-modal data (time-seriesFM, weatherFM, LLM):</p> <ul style="list-style-type: none"> • Load & renewable generation forecasting 	<p>Optimal Power Flow: reconstruction under constraints and objective:</p> <ul style="list-style-type: none"> • Loss encoded constraints & objective • Reinforcement learning 	<p>Model to deal with spatio-temporal data enabling:</p> <ul style="list-style-type: none"> • Instability localization & anomaly detection from PMU data • Disaggregation (super-resolution) from AIM and PMU covariates 	<p>Scenario analysis by chain-of-thought, agentic workflows:</p> <ul style="list-style-type: none"> • Remedial-action • Expansion planning • Explainable AI

Let's scale the Models
and enjoy the Workshop!